

VISIT...

LANZAROTE
Caliente.COM

KIRK-OTHMER

**ENCYCLOPEDIA OF
CHEMICAL
TECHNOLOGY**

EDITING - by **WATCHER**

VOLUME 18

FOURTH EDITION

**PAPER TO
PIGMENT DISPERSIONS**

CONTENT INDEX (vol 18)

(with hyperlinks)

Edited by <http://www.watcherworld.narod.ru/>

PAPER	1
PAPERMAKING ADDITIVES	15
PATENTS AND TRADE SECRETS	25
PATENTS, LITERATURE	42
PERCHLORIC ACID AND PERCHLORATES	64
PERFUMES	71
PEROXIDES AND PEROXIDE COMPOUNDS	89
- INORGANIC PEROXIDES	90
- ORGANIC PEROXIDES	101
PESTICIDES	141
PETROLEUM	158
- NOMENCLATURE IN THE PETROLEUM INDUSTRY	158
- ORIGIN OF PETROLEUM	161
- COMPOSITION	164
- DRILLING FLUIDS	174
- ENHANCED OIL RECOVERY	188
- REFINERY PROCESSES, SURVEY	200
- PETROLEUM RESOURCES	217
PHARMACEUTICALS	223
PHARMACEUTICALS, CHIRAL	237
PHARMACODYNAMICS	267
PHENOL	286
PHENOLIC RESINS	292
PHOSGENE	311
PHOSPHINE AND ITS DERIVATIVES	317
PHOSPHORIC ACIDS AND PHOSPHATES	323
PHOSPHORUS	347
PHOSPHORUS COMPOUNDS	356
PHOTOCHEMICAL TECHNOLOGY	387
- SURVEY	388
- PHOTOCATALYSIS	398
PHOTOCONDUCTIVE POLYMERS	407
PHOTODETECTORS	419
PHOTOGRAPHY	440
PHOTOVOLTAIC CELLS	466
PHTHALIC ACIDS AND OTHER BENZENEPOLYCARBOXYLIC ACIDS	478
PHTHALOCYANINE COMPOUNDS	504
PIGMENT DISPERSIONS	511

PAPER

Paper consists of sheet materials that are comprised of bonded small discrete fibers. The fibers usually are cellulosic in nature and are held together by hydrogen bonds (see CELLULOSE). The fibers are formed into a sheet on a fine screen from a dilute water suspension. The word paper is derived from papyrus, a sheet made in ancient times by pressing together very thin strips of an Egyptian reed (*Cyperus papyrus*) (1). The early use of paper was as a writing medium which replaced clay tablets, stone, parchment, and papyrus sheets. Papyrus sheets are not considered paper because the individual vegetable fibers are not separated and then reformed into the sheet. Paper apparently originated in China in 105 AD and was made from flax and hemp or bark fibers of certain trees (2). The manufacture of paper from bark and bamboo spread from China to Japan where its manufacture began in ca 610. Papermaking from flax and hemp spread through Central Asia, the Middle East, and eventually into Europe. Spain and Italy were the early European papermaking centers. The first European paper was made in Spain in 1150, in France by 1189, in Germany by 1320, and in England by 1494. Papermaking was introduced in the United States in ca 1700 by William Rittenhouse in Philadelphia; a history of U.S. papermaking is available (3).

Paper is made in a wide variety of types and grades to serve many functions. Writing and printing papers constitute ca 30% of the total production. The balance, except for tissue and toweling, is used primarily for packaging (qv). Paperboard differs from paper in that it generally is thicker, heavier, and less flexible than conventional paper.

More than 95% of the base material used in paper and board manufacture (4) is fibrous. Whereas a large (90%) percentage originates from wood (qv), the filler contents of some grades of paper approach 30%. Many tree species encompassing both hardwood and softwood are used to produce pulp. In addition to the large number of wood types, there are many different manufacturing processes involved in the conversion of wood to pulp. These range from mechanical processes, by which only mechanical energy is used to separate the fiber from the wood matrix, to chemical processes, by which the bonding material, ie, lignin (qv), is removed chemically. Many combinations of mechanical and chemical methods also are employed (see PULP). Pulp properties are determined by the raw material and manufacturing process, and must be matched to the needs of the final paper product.

In mechanical pulps, the fibers are separated by mechanical energy. Because there is no chemical removal of wood components, the process results in a high (95%) pulp yield. The chemical composition of mechanical pulps is similar to those of native wood, ie, they contain significant amounts of lignin and hemicellulose (qv) in addition to the basic cellulose component of the fiber. In the stone groundwood process, fibers are separated from the wood by grinding logs against revolving stone wheels. The resultant fibers are fractured and much fibillar debris is generated. These pulps are used where opacity and good printability are needed. The presence of large amounts of lignin, however, results in lower interfiber bonding and tensile strength, and poor light stability. Such pulps are bleached with sequences that maintain a high pulp yield and do not remove lignin. Typical bleaching chemicals are either alkaline hydrogen peroxide or sodium hydrosulfite (dithionite). Brightness levels of 80% usually can be achieved.

Newer methods of mechanical pulp production involve disk refiners to produce pulp from wood chips. Chips are passed between closely spaced revolving disks and the fiber is broken free from the wood material. The pulps generally contain less debris and longer fibers than stone groundwood pulp. Refiner mechanical pulp (RMP) is produced at atmospheric pressure from a disk refiner; it is the oldest of the refiner processes. In the thermomechanical pulp (TMP) process, chips are steamed at 120°C prior to fiberization in a pressurized disk refiner. Compared to stone groundwood pulp, TMP is comprised of longer, less damaged fibers and less debris, which results in improved strength but some opacity loss. Thermomechanical pulps can be used to reduce or eliminate chemical pulps in many blends. Softwood is the preferred raw material to produce optimum strength in the product.

Certain chemical treatments can be employed during the TMP process to achieve improved strength. Sodium sulfite and hydrogen peroxide have been used either for chip pre- or post-treatment of the TMP pulp; such pulp is called chemithermomechanical pulp (CTMP). The strength improvements, which may be 50%, are obtained at some sacrifice to yield and opacity. The yields of mechanical pulps are 90–95%; the lower yields are associated with chemical treatment. No principal commercial pulps are produced in the next lower yield range, ie, 80–90%.

The next significant class of pulps, semichemical pulps, generally is characterized by a yield of less than 80%; significant amounts of material are removed by chemical action. Semichemical pulps are produced by mild chemical digestion of chips prior to reduction to pulp in a disk refiner. Yields are 70–80%, and the pulp contains a lower lignin and hemicellulose content than the wood from which it was derived. The main use for this product is in corrugated media, in which the stiffness resulting from the lignin and hemicellulose components is a product advantage. Hardwoods usually are the base woods for semichemical pulps. Use of the neutral sulfite process is declining because of environmental problems resulting from the lack of a suitable recovery system. Newer processes, which are based on sodium carbonate–sodium sulfide, ie, the green liquor process, or sodium hydroxide–sodium carbonate, are replacing the neutral sulfite sequence. Compared to the neutral sulfite pulps, these alkaline semichemical pulps are darker in color and have slightly higher lignin contents.

Chemical pulps have greatly reduced lignin and hemicellulose contents compared to the native wood, as these components are dissolved during chemical digestion. Because the lignin is removed, much less mechanical energy is needed to separate the fibers from the wood matrix, and the resulting pulp fibers are undamaged and strong. Chemical pulps are used principally for strength and performance in a variety of paper and paperboard products.

The principal process for producing chemical pulps is the kraft process: mixtures of sodium sulfide and sodium hydroxide are the pulping chemicals, and yields are 46–56%. The higher yield pulps contain about 10% lignin. They are used in bags or linerboard where strength is important. The lower yield pulps typically are bleached to remove virtually all lignin and to produce high brightness (90%+). These pulps are used where permanence and whiteness are needed in addition to strength. Bleaching is a multistage process. Elemental chlorine was commonly used as a bleaching agent (see BLEACHING AGENTS, PULP AND PAPER), but the environmental impact of effluents from chlorine bleacheries which contained dioxins led to replacement by chlorine dioxide (see CHLORINE OXYGEN ACIDS AND SALTS). In order to control the formation of dioxins, there has been a trend toward extended cooking and/or oxygen delignification, followed by chlorine dioxide using oxygen augmentation of extraction stages or occasionally oxygen and peroxide augmentation. The use of oxygen compounds, eg, gaseous oxygen, ozone (qv), and hydrogen peroxide (qv), is becoming more important.

Reclaimed fiber accounts for ca 40% of the total fiber used in the United States. A variety of sequences are used to disperse and clean the waste fiber. Emphasis is on mechanical screening and cleaning, washing, and floatation. The properties of these pulps depend largely on the input raw material and generally are lower in strength and brightness than a comparable virgin pulp. Typical yields for de-inking processes range from 60 to 80%.

Nonwood fibers are used in relatively small volumes. Examples of nonwood pulps and products include cotton linters for writing paper and filters, bagasse for corrugated media, esparto for filter paper, or Manila hemp for tea bags. Synthetic pulps which are based on such materials as glass (qv) and polyolefins also are used (see OLEFIN POLYMERS). These pulps are relatively expensive and usually are used in blends with wood pulps where they contribute a property such as tear resistance, stiffness, or wet strength which is needed to meet a specific product requirement.

Physical Properties

Most properties of paper depend on direction, ie, the machine, cross-machine, and thickness directions. For example, strength is greater if measured in the machine direction, ie, the direction of manufacture, than in the cross-machine direction. For paper made on a Fourdrinier paper machine, the ratio of the two values varies from about 1.5 to 2.5. An even greater anisotropy is observed if either of the in-plane values is compared to the out-of-plane strength. Paper is quite weak in the thickness direction; it may be considered an orthotropic material, ie, one possessing three mutually perpendicular symmetry planes (5). There are several reasons for this anisotropy. Wood pulp fibers are long, slender, and usually ribbon-like after paper formation, rather than circular in cross section. During the deposition of the slurry onto the wire, the fibers tend to line in the direction of the moving wire, and the extent to which they do this depends largely on the ratio of the jet and wire speeds. The sheet tends to become stronger in the machine direction as more fibers line up in the same direction. Another important factor is the tendency for paper to become stronger if it is dried under restraint, ie, prevented from shrinking

during drying. Machine-made paper is dried by being passed through an array of drum dryers. The paper is under tension in the machine direction during this process.

Because the fibers generally are anisotropic, they tend to be deposited on the wire in layers under shear. There is little tendency for fibers to be oriented in an out-of-plane direction, except for small undulations where one fiber crosses or passes beneath another. The layered structure results in the different properties measured in the thickness direction as compared to those measured in the in-plane direction. The orthotropic behavior of paper is observed in most paper properties and especially in the electrical and mechanical properties.

The *basis weight*, *W*, is the mass in g/m² (TAPPI T410). It can also be expressed as pounds of a ream of 500 sheets of a given size, but the sheet sizes are not the same for all kinds of paper. Typical sizes are 43.2 × 55.9 cm for fine papers, 61.0 × 91.4 cm for newsprint, and 63.5 × 96.5 cm for several book papers. The most common designation is pounds per 3000 square feet (1.62 g/m²) for paper. The basis weight of board usually is expressed as pounds per thousand square feet (kg/205 m²). For example, the material comprising a 69-lb (31-kg) linerboard weighs 69 lb/1000 ft² (337 g/m²). Typical basis weights are tissue and toweling, 16–57 g/m²; newsprint, 49 g/m²; grocery bag, 49–98 g/m²; fine papers, 60–150 g/m²; kraft linerboard, 127–439 g/m²; and folding boxboard, 195–586 g/m².

The caliper is the thickness in μm of a single sheet measured under specified conditions (TAPPI T411). Calipers for a number of common paper and board grades are capacitor tissue, 7.6 μm; facial tissue, 65 μm; newsprint, 85 μm; offset bond, 100 μm; linerboard, 230–640 μm; and book cover, 770–7600 μm.

The tensile strength is the force per unit width parallel to the plane of the sheet that is required to produce failure in a specimen of specified width and length under specified conditions of loading (TAPPI T404). The strength of paper also is expressed in terms of a breaking length, ie, the length of paper that can be supported by one end without breaking. Breaking lengths for typical papers are from ca 2 km for newsprint to 12 km for linerboards. The values for the stronger papers compare favorably with other engineering materials. For example, breaking lengths for aluminum are ca 20–25 km.

Stretch is the extension or strain resulting from the application of a tensile load applied under specified conditions (TAPPI T457). The numerical result usually is expressed as a percentage of elongation per original length and includes the elastic and the inelastic extensibility of the paper. Stretch is greatest in the cross-machine direction except for creped grades. It is becoming more common to evaluate the elongation as a continuous function of the applied load. The initial slope of the load–elongation curve, ie, load–width versus strain, defines the modulus of elasticity, *E*, in the machine or cross-machine directions. By this definition, *E* describes units of load per width.

The bursting strength is the hydrostatic pressure required to rupture a specimen when it is tested in a specified instrument under specified conditions. It is the pressure required to produce rupture of a circular area of the paper (30.5-mm dia) when the pressure is applied at a controlled rate (TAPPI T403). It is related to tensile strength and extensibility and is used extensively throughout the industry for packaging and container grades.

Tearing strength, or the internal tearing resistance, is the average force required to tear a single sheet of paper under standardized conditions by which the specimen is cut prior to tearing (TAPPI T414). Internal tearing resistance should be distinguished from initial or edge-tearing resistance.

Stiffness is related to bending resistance. It is most commonly measured by determining the force required to produce a given deflection or by measuring the deflection produced by a given load when the paper specimen is supported rigidly at one end and the deflecting force is applied at the free end. A fundamental measure of stiffness is the flexural rigidity of the sample which is the product of the modulus of elasticity and the second moment of the cross section, *I*. If evaluated in accordance with TAPPI T451, stiffness values are proportional to *EI/W*. All other factors being the same, the stiffness of paper varies with the cube of the thickness and directly with the modulus of elasticity.

Folding endurance refers to the number of folds a paper can withstand before failure when tested (TAPPI T423).

Moisture content is determined by drying the sample at 100–105°C, until no change in weight is observed (TAPPI T412). Unless otherwise specified, the difference in the before and after drying weights is expressed as a percentage of the original weight of the sample. Water resistance refers to that property of a sheet that resists passage of liquid water into or through the sheet. The tests usually are designed to simulate use conditions; consequently, there are several different test methods. The Hercules size test (HST) is commonly performed in paper mills in order to gauge the penetration of water at the size press, which, if excessive, can lead to the paper web breaking and poor machine runnability (6). The Cobb water absorption test measures the total amount of water absorbed by paper from one side in one minute (6). For prediction of the rate of penetration of various liquids into paper under dynamic conditions, such as experienced in printing, gluing, or coating, a Bristow wheel test is often employed (6).

Water-vapor permeability refers to a specific permability of the paper to water vapor. Two common gravimetric methods for evaluating this property are available (TAPPI T448 and T464). The tests differ in temperature and vapor pressure differences which cause permeation. Permeability usually is reported in grams of vapor permeating one square meter of paper per 24 hours. Because of the unusually high affinity of cellulose for water and water vapor, water-vapor permeability generally does not correlate with permeability to other vapors and gases.

The common optical properties of paper are brightness, color, opacity, transparency, and gloss. Brightness is the reflectivity of a sheet of pulp or paper for blue light, ie, ca 457 nm (TAPPI T452). The reflectivity is the reflectance for an infinitely thick sample. Color is measured by evaluating the spectral reflectivity (TAPPI T442). Opacity relates to that property of a sheet which prevents dark objects in contact with the back side of the sheet from being seen. It usually is evaluated by contrast ratio, which is the ratio of the diffuse reflectance of the sheet when backed by a black body to that of the sheet when backed by a white body of given absolute reflectance value (TAPPI T425). Transparency is that property of a paper by which it transmits light, so that objects can be seen through the paper; transparency ratio is a measure of transparency as judged when a space separates the specimen and the object being viewed. Gloss is the ability of the surface to reflect light specularly. There are numerous definitions of gloss as it relates to appearance criteria for paper, eg, specular gloss (TAPPI T480) and low angle gloss, which often is used as a smoothness test for linerboard (TAPPI UM (useful method) 558).

Chemical Properties

The chemical composition of paper is determined by the types of fibers used and by any nonfibrous substances incorporated in or applied to the paper during the papermaking or subsequent converting operations. Paper usually is made from cellulose fibers obtained from the pulping of wood. Occasionally, synthetic fibers and cellulose fibers from other plant sources are used. Paper properties affected directly by the fibers' chemical composition include color, opacity, strength, permanence, and electrical properties. Development of interfiber bonding during papermaking also is strongly influenced by the composition of the fibers. Groundwood pulp, which contains some residual lignin, is used in newsprint and in some book and absorbent papers which do not require a highly bonded structure. Lignin contains chromophores which can revert after bleaching to yellow-brown in the presence of ultraviolet light. Thus, lightfastness of paper is usually assessed in a Fadeometer which subjects the paper to intense ultraviolet light while maintaining the sample at room temperature (see FIBERS).

Mechanical permanence depends principally on the pH of paper. Whereas lignin-containing pulps tend to yellow with aging, experimental evidence demonstrates that lignin does not cause strength loss with aging. Papers that are manufactured under neutral or mildly alkaline conditions can maintain their physical properties for hundreds of years. In order to prevent the acidification of papers by the absorption of atmospheric pollutants such as sulfur dioxide and nitrous oxide, it is also important that the paper contain an alkaline reserve, usually in the form of calcium carbonate used as a filler (6). Fine paper is made predominantly under alkaline conditions in order to take advantage of low cost bright minerals, such as ground and precipitated calcium carbonates, as fillers (qv). In the United States, the transition to alkaline papermaking has been dramatic in the 1990s.

Hemicelluloses in chemical pulps contribute to bonding; therefore, pulps containing hemicelluloses are used for wrapping papers and other grades which require bonding for strength, and in glassine which requires bonding for transparency. In most papers, the chemical composition largely reflects those nonfibrous materials that were added to the paper to achieve the desired physical, optical, or electrical properties (see HEMICELLULOSE; PAPERMAKING ADDITIVES). Examples of chemicals and resultant properties are dyes and optical brighteners to enhance appearance, resins to impart wet strength, rosin or starch size to reduce penetration of aqueous liquids, pigment coatings to provide a smooth surface for printing, mineral fillers to increase opacity, polymers applied by saturation or extrusion to impart mechanical or barrier properties, and cationic polyelectrolytes and resistive polymers used in the interior and on the surface, respectively, of papers for dielectric recording. The performance of a limited number of papers depends on chemical reactions of noncellulosic

additives. Specialized paper coatings in which chemical reactions occur at the time of use are essential in photographic, thermal, and carbonless copy papers (see MICROENCAPSULATION; PHOTOGRAPHY). Strips of paper saturated with color-forming reagents permit rapid, inexpensive urinalyses (see AUTOMATED INSTRUMENTATION, CLINICAL CHEMISTRY). Phosphates or halogenated compounds are incorporated in papers to promote flame retardancy (see FLAME RETARDANTS). For acceptable performance, some grades of paper should not contain certain chemical species. Papers used for electrical insulation must be free of electrolytes, and papers used for permanent documents must be low in acidity. Reducible sulfur compounds, ie, sulfide, elemental sulfur, and thiosulfate, should not be present in papers that receive metallic coatings or in antitarnish papers that are used for wrapping polished silver or steel items. Chemical substances that unintentionally could become a component of food must not be included in papers for food contact applications unless the chemicals are approved by the U.S. FDA as indirect food additives (qv).

Manufacture and Processing

Stock Preparation. Stock preparation denotes the several operations that must be undertaken in order to prepare the furnish from which paper is made (7–9). During the stock preparation steps, papermaking pulps are most conveniently handled as aqueous slurries so that they can be conveyed, measured, subjected to desired mechanical treatments, and mixed with nonfibrous additives before being delivered to the paper machine. In the case of adjacent pulping and papermaking operations, pulps usually are delivered to the paper mill in slush form directly from the pulping operation. Purchased pulps and waste paper are received as dry sheets (laps) and must be slushed before use. The objectives of slushing are to separate the fibers and disperse them in water with a minimum of mechanical work so as not to alter the fiber properties. Slushing is accomplished in several types of apparatus, eg, the Hydrapulper (Fig. 1).

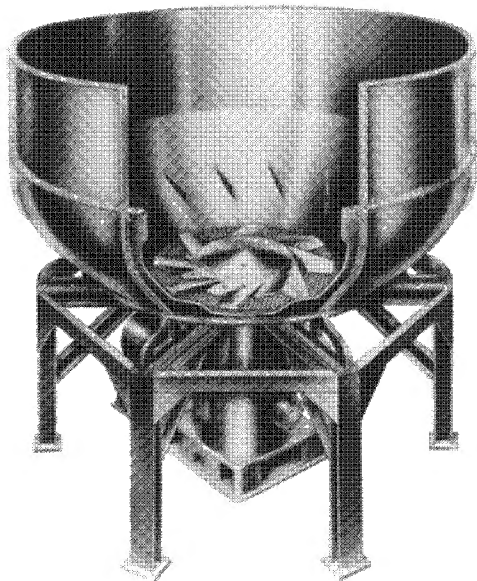


Fig. 1. The Hydrapulper.

Courtesy of Black Clawson Co.

Beating and Refining. Virtually all pulps are subjected to certain mechanical actions before being formed into a paper sheet. Such treatments are used to improve the strength and other physical properties of the finished sheet and influence the behavior of the system during the sheet-forming and drying steps. During refining, the cellulose fibers are swollen, cut, macerated, and fibrillated. The most desirable action is the development of internal fibrillation, which makes the wet fiber more compliant or flexible. This flexibility enhances the number of interfiber contacts during formation of the paper, and bonding during subsequent pressing and drying operations. Although refining markedly alters the physical properties, no significant chemical changes occur. A sheet formed from an unbeaten pulp has low density and is rather soft and weak. If the same pulp is well beaten, however, the resultant paper is much more dense, hard, and strong. If taken to the extreme, beating produces very dense, translucent, glassine-type sheets. Refining greatly increases the wet specific surface of pulp fibers, swollen specific volume, and fiber flexibility. Although hydration in the chemical sense does not occur, the affinity for water is enhanced. Because of the unique cellulose–water relationship, these changes significantly increase the ability of the fibers to bond when dried from a water suspension and, therefore, enhance the strength of the sheet. Optical micrographs of a softwood kraft pulp before and after several periods of extensive beating are given in Figure 2. In Figure 3, the tensile strength, bursting strength, and tearing resistance are shown as functions of beating time for a softwood kraft pulp. It generally is true that, within the commercial range, beating increases tensile strength, bursting strength, folding endurance, and sheet density, whereas it reduces tearing resistance.

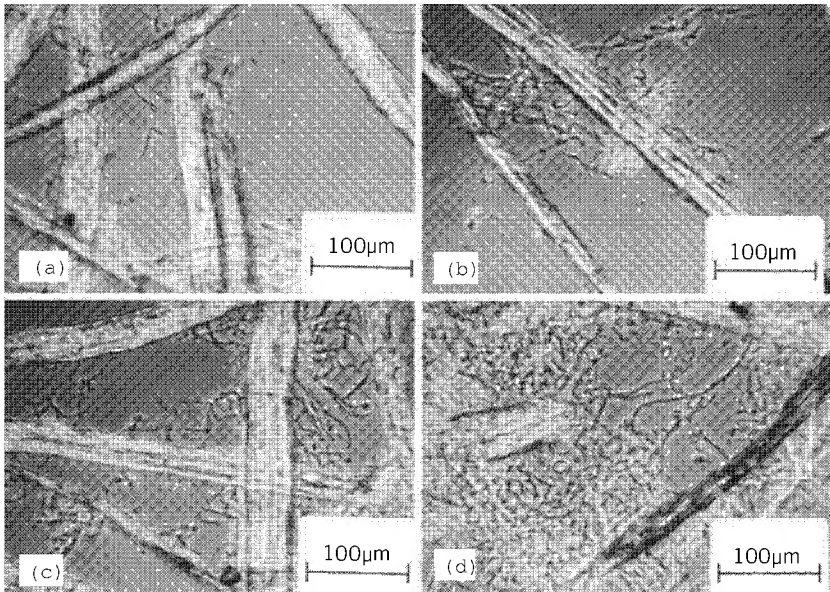


Fig. 2. Optical micrographs of a softwood sulfite pulp (a) before beating and (b–d) after beating in a valley beater for different periods of time. Beating

time increases from (b) to (d).

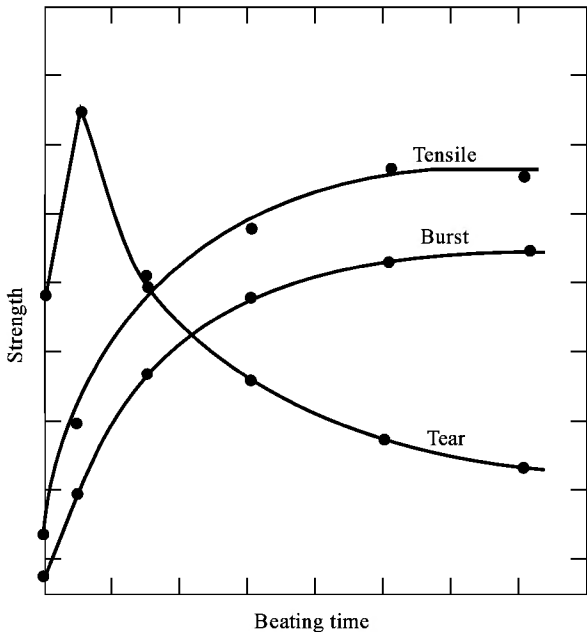


Fig. 3. Beating curves for a softwood kraft pulp.

Batch systems have been replaced largely with continuous, pump-through equipment. Most batch systems are refinements of the beater developed in Holland in ca 1690. This hollander beater consists of an elongated tub with a central dividing partition (midfeather) which extends along the major axis to within a few meters of each end. A cylindrical beater roll is mounted on one side and knives are placed around the circumference parallel to the roll axis. The roll is mounted over a bedplate, which also contains a set of knives. Circulating stock passes between the roll and the bedplate, and the severity of beating is controlled by adjusting the load of one against the other. Most modifications of the original hollander represent attempts to increase capacity, improve stock circulation, and save power. The first successful continuous refiner was the Jordan, developed in ca 1860. The Jordan refiner consists of a stationary conical shell fitted with knives or bars on the inside, and a conical plug which fits inside the shell and contains bars on its surface. The plug rotates, and the pressure between the bars on the shell and the plug is regulated by longitudinal movement of the plug. The pulp slurry is fed into the small end of the Jordan refiner and is discharged at the large end.

A variety of conical refiners were developed that involved modifications of the Jordan refiner. The Hydrafiner, for example, has a short, low taper, high speed rotor, and wide bars (10). The stock is driven through the refiner by an impeller which is fitted to the small inlet end of the rotor shaft. The Claflin refiner has a very short, high taper plug. Vanes that fit on the wide end of the plug draw the stock through the unit (11).

The disk refiner, a newer development used for fine paper grades, includes one or two rotary disks and two or four working surfaces. The surfaces are pressed together uniformly by hydraulic pressure and guided by high precision, heavy-duty bearing systems. Stock usually is fed through the center of one plate and leaves between the plates at their circumferences.

A considerable variation in the properties of the refined stock may be realized through control of the refiner operating variables. For example, high stock consistency, dull beater or refiner bars, or low pressure or excessively wide separation between the two sets of bars may yield a mild refining action resulting in accentuated fibrillation and swelling. Low stock consistency, sharp bars, and high pressure between elements can result in severe refining and severe fiber length reduction with subsequent negative effects on strength.

Filling and Loading. Materials, eg, mineral pigments for filling and loading, are added to the pulp slurry to make the papermaking furnish (see PAPERMAKING ADDITIVES). Mineral fillers are used in varying amounts, depending on the grade of paper, and may comprise 2–40 wt % of the final sheet. Fillers can improve brightness, opacity, softness, smoothness, and ink hold-out. They almost invariably reduce the degree of sizing and the strength of the sheet. The brightness, particle size, and refractive index of fillers influence the optical properties of the finished sheet, and the particle size and specific gravity are important in regard to the filler retention during sheet formation. Most commercial fillers are essentially insoluble in water under the conditions of use.

Kaolin or China clay is used both as a filler material and as a coating pigment. It is a low cost, naturally occurring, hydrated aluminum silicate with widespread application. Titanium dioxide probably is the most desirable pigment for opacity improvement, and its use is increasing, particularly in fine papers. Two forms of titanium dioxide are used, ie, anatase and rutile. Calcium carbonate is used particularly in book and cigarette papers, as well as in coated and uncoated free sheets and in medium weight coated mechanical pulp-containing grades. Formed at the mill site by reaction of carbon dioxide from flue gas and lime, various crystal forms and particle sizes can be obtained by varying the conditions for precipitation. Calcium carbonate filler can also be obtained from natural sources such as limestone or chalk by grinding (see LIME AND LIMESTONE).

Calcium carbonate is not used in papers that are sized in an acid furnish because of its solubility and resultant foam problems. However, the conversion of the fine paper industry to alkaline papermaking since the 1980s has made calcium carbonates the minerals of choice because they are generally brighter than kaolin. Conversion from acid to alkaline papermaking is attractive because it results in stronger paper by virtue of greater swelling and conformability of fibers in an alkaline medium. Thus, it is possible to introduce more filler at the same strength level, thereby obviating the need for expensive titanium dioxide to gain opacity and brightness. Some properties of important fillers are listed in Table 1.

Table 1. Properties of Paper Fillers

Filler	Specific gravity	Refractive index	Particle size, μm	Brightness, %
clay	2.5–2.8	1.55	0.5–1.0	80–85
titanium dioxide, anatase	3.9	2.55	0.3	98–99
rutile	4.2	2.70	0.3	98–99
calcium carbonate	2.6–2.8	1.56	0.2–0.4	95–97
zinc sulfide	4.0	2.37	0.3	96–98
talc	2.8	1.57	1–10	70–90
synthetic silicates	2.1	1.55	0.1–4.0	93–94

The retention of fillers in the sheet during the forming process is important. Both hydrodynamic mechanisms and colloidal or coflocculation phenomena are significant in determining filler retention (7). Polymeric retention aids are used to bridge between filler particles and fibers. Talc is sometimes used with mechanical pulp furnishes in order to reduce the deposition of pitch-like materials onto paper machinery.

Sizing. Sizing is the process of adding materials to the paper in order to render the sheet more resistant to penetration by liquids, particularly water. Unsized or waterleaf paper freely absorbs liquids. Writing and wrapping papers are typical sized sheets, as contrasted with blotting paper and facial tissue which usually are unsized. In the past, rosin with aluminum sulfate (alum) mordant was commonly added to the pulp furnish to increase the hydrophobicity of paper made under acid conditions. In the 1990s, synthetic polymers such as alkyl ketene dimer (AKD) and alkyl succinic anhydride (ASA) are commonly used as sizing agents at the wet end of the paper machine. Alternatively, sizing can be added to the surfaces of the sheet in the dryer section of the paper machine. The size press can be configured to pass the paper web through a size solution or over a roll that has been wetted with a size solution. Generally, starch (qv) is applied with the surface size in order to increase the surface integrity of papers, thereby decreasing linting, dusting, and picking in uses such as printing and photocopying.

Rosin, extracted from pine stumps or fractionated from crude tall oil (qv), is a by-product of the kraft pulping process. It is partially saponified with caustic soda and is often processed to yield a thick paste of 70–80 wt % solids. Dry rosin sizes also are available. At the paper mill, the paste is diluted to ca 3 wt % solids with hot water and vigorous agitation. The solution is added to the stock (0.5–3.0 wt % size based on dry fiber) usually before but sometimes simultaneously with one to three times as much aluminum sulfate, which precipitates the rosin on the fibers as flocculated particles. The pH after the addition of the alum is critical and should be in the 4.5–5.5 range for the mordant effect. In systems of higher pH, sodium aluminate also may be used to precipitate the rosin size. Internal sizing of papers and paperboard under alkaline papermaking conditions (pH 7.0–8.5) is achieved using synthetic sizes having trade names such as Aquapel (12), Hercon (13), and Fibran (14). Surface sizes vary according to the use of the paper. For example, polyurethane provides effective hydrophobicity, but cannot be used on copy papers to which powdered toner must adhere. Styrene maleic anhydride and styrene acrylic acid are commonly used for printing and copy papers.

Coloring. The color of most paper and paperboards made from bleached pulps is achieved by the addition of dyes (qv). White papers frequently are treated with small amounts of blue dye to achieve a whiter visual appearance. By far the largest proportion of dyes is added during stock preparation, although color is sometimes corrected at the size press or by applying a dye solution during calendering. Some grades, such as file folder, are surface tinted at the size press. Water-soluble synthetic organic dyestuffs are the principal paper-coloring materials. Some coloring is by water-insoluble but water-dispersible pigments, eg, carbon black, and sulfur colors. The properties of basic, acid, and direct dyes are summarized in Table 2. Within each group there are wide variations and exceptions from the listed generalizations.

Table 2. Comparison of Basic, Acid, and Direct Dyes

Parameter	Basic dyes	Acid dyes	Direct dyes
cation	dye ion	Na ⁺ , K ⁺ , NH ₄ ⁺	Na ⁺
anion	Cl ⁻ , SO ₄ ²⁻ , NO ₃ ⁻	dye ion	dye ion
tinctorial strength	high	lower	lower
brilliancy	high	high	lower
lightfastness	poor	generally good	good
acid fastness	poor	poor	variable
alkali fastness	poor	poor	variable
waterbleed fastness	generally good	generally good	generally good
solubility	good	high	lower
affinity	strong for unbleached lignified fibers ^a ; no mordant necessary	none for cellulose; a mordant, eg, size and alum, is necessary	very strong for bleached or unbleached cellulose

^a Mottling occurs in mixed furnishes.

The kind of fiber and the degree to which it has been refined are important factors in paper dyeing. The undyed color of the pulp and the varying affinity for the same or different dyes, both from fiber to fiber within a pulp and between different pulps, are some of the variables which necessitate continual adjustment of dyeing techniques. The amount and kind of refining a pulp has received affects the pulp's optical properties and, therefore, the color effect of a given dye. Generally, refining deepens the shade from a given application of a water-soluble dye but does not change the amount of dye that is retained. Many types of dyes are absorbed strongly by a wide range of fillers. The two-sidedness of sheets, which results from loss of fines and filler on the side in contact with drainage elements on the paper machine (ie, bottom, or wire side on a single wire paper machine), contributes to two-sidedness of the color, an effect that also may result from or be enhanced by contact with heated dryer surfaces. There is a tendency for pigment colors to be concentrated on the top side of the sheet. In a complex system containing fibers, fillers, size, and dye, much colloidal activity is possible, particularly when extraneous unknown ions or particles are present. An optimum order for the addition of the filler, size, alum, and color is difficult to predict.

Other Beater Additives. Beater adhesives are employed widely to enhance fiber-to-fiber bonding. Starches are used in the greatest tonnage. Natural gums (qv), eg, guar and locust bean, also are used as are modified celluloses, eg, the carboxymethyl and hydroxyethyl derivatives (see CELLULOSE ESTERS). Urea–formaldehyde and melamine–formaldehyde polymers provide wet strength to the finished paper sheet (see AMINO RESINS AND PLASTICS). Polyamide epichlorohydrine and dialdehyde starch are also wet-strength resins. The former is used in neutral papers. Other natural and synthetic materials are used to alter the paper properties and to influence the behavior of the system during sheet forming and drying.

Sheet Forming, Pressing, and Drying

Continuous sheet forming and drying came into use in ca 1800. The equipment was of two types: the cylinder machine and the Fourdrinier machine. In the former, a wire-covered cylinder is mounted in a vat containing the fiber slurry. As the cylinder revolves, water drains inward through the screen and the paper web is formed on the outside. The wet web is removed at the top of the cylinder, passes through press rolls for water removal, and then passes into steam-heated, cylindrical drying drums. The Fourdrinier is more complex and basically consists of a long continuous wire screen which is supported by various devices that improve drainage. The fiber slurry, which is introduced at one end through a headbox and slice, loses water as it progresses down the wire, thereby forming the sheet. It then passes to presses and dryers as in the cylinder machine. A Fourdrinier paper machine is shown diagrammatically in Figure 4.

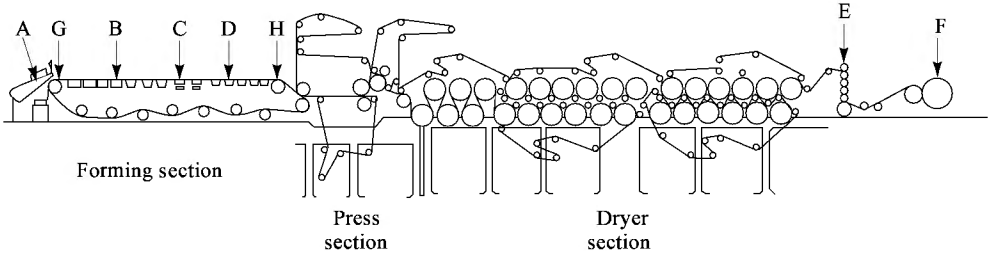


Fig. 4. Fourdrinier paper machine: A, headbox; B, Fourdrinier wet end with foil boxes; C, wet and D, dry suction boxes, pickup, and closed transfer of web through the press and dryer sections; E, calender; F, reel; G, breast roll; and H, couch roll.

Courtesy of Beloit Corp.

Continuous paper machines have undergone extensive mechanical developments since the 1950s, although the principles employed have changed little. Cylinder machines still are operated and involve multiples of five to seven cylinders; they are used to produce heavy multi-ply boards. Fourdriniers are common in the industry and can be used to produce virtually any grade of paper or paperboard. They vary from 1 to 10 meters in width and, including the press and dryer sections, may be more than 200 m long.

Subsequent to stock preparation and proper dilution, the paper furnish usually is fed to the paper machine through one or more screens or other devices to remove dirt and fiber bundles. It then enters a flow spreader which provides a uniform flowing stream and which is the width of the paper machine. The flow spreader, or manifold, discharges the slurry into a headbox, where fiber flocculation is minimized by microturbulence and where the proper pressure head is provided to cause the slurry to flow at the proper velocity through the slice and onto the moving Fourdrinier wire.

The Fourdrinier wire is mounted over the breast roll at the intake end and at the couch roll at the discharge end. Between the two rolls, it is supported for the most part by foils and suction boxes. Foils (Fig. 5) are wing-shaped elements that support the wire and induce a vacuum at the downstream nip. Foil geometry can be varied to provide optimum conditions. After passing over the foils, the wire and sheet pass over suction boxes where more water is removed. Most machines also include a suction couch roll for additional water removal. Machine speeds vary chiefly because of limitations imposed by the various products but also because of differences in production equipment. Heavy paperboards require a long drying time, and machine speeds are 50–250 m/min. Very dense papers, eg, glassine and greaseproof and condenser tissue, are difficult to dewater in the forming and press sections; speeds range from 20 to 300 m/min, depending on the product. Brown grades, eg, paper bags and linerboard, are produced at 200–1000 m/min, depending on basis weight and the site of the paper machine. With the advent of the suction pickup, which closes the draw between the forming and press sections, speeds of newsprint machines increased from 400 to 800 m/min. Closing the transfer of the sheet through the entire press section and increasing the dryness at the first open draw into the dryer section from 35–38% to 41–44% combined with careful designs of the web path through the dryer section has increased newsprint machine speeds to ca 1500 m/min. The majority of machines operate at 800–1200 m/min. Drying capacity restraints and difficulties in reeling the product limit modern tissue machine speeds to 1500–1800 m/min. Most tissue machines operate at lower speeds. Novel designs for web handling, reeling, and roll change will permit tissue machine speeds of up to ca 2000 m/min in continuous operation.

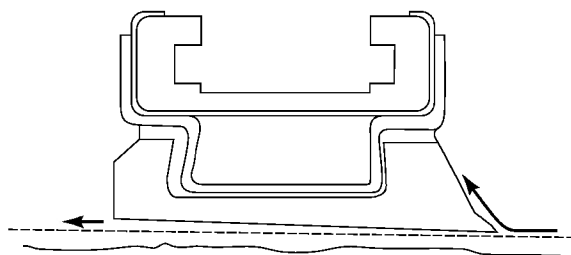


Fig. 5. Dewatering foil. Water from the preceding unit is doctored off. The diverging wedge on the downstream side of the foil vacuums water out of the slurry onto the wire.

Virtually all newly installed paper machines are twin-wire formers because they offer more stable high speed operation and better control of forming and dewatering conditions. Many large Fourdrinier wet ends have been retrofitted with top wire units to achieve similar advantages under high speed operation. These retrofits have been particularly popular for lightweight sheets, eg, tissue, towel, and newsprint. However, twin-wire formers also are operated successfully on fine paper grades, corrugated media, and linerboard grades.

In twin-wire formers, the water is drained from the slurry by pressure rather than vacuum. The two wires, with the slurry between, are wrapped around a cylinder or a set of supporting bars or foils. The tension in the outer wire results in a pressure that is transmitted through the slurry to the supporting structure. The pressurized slurry drains through one or both of the wires. The Bel Baie II is an example of a twin-wire former (Fig. 6a). It is used extensively for newsprint and fine paper grades as well as for the lighter weight linerboards. Dewatering occurs through both wires; thus a high drainage capacity is achieved. In a typical roll-type twin-wire tissue former (Fig. 6b) drainage is single sided and limited to low basis weights. The tissue former is sufficient for drainage at very high speeds, eg, ≥ 2100 m/min for thin tissue.

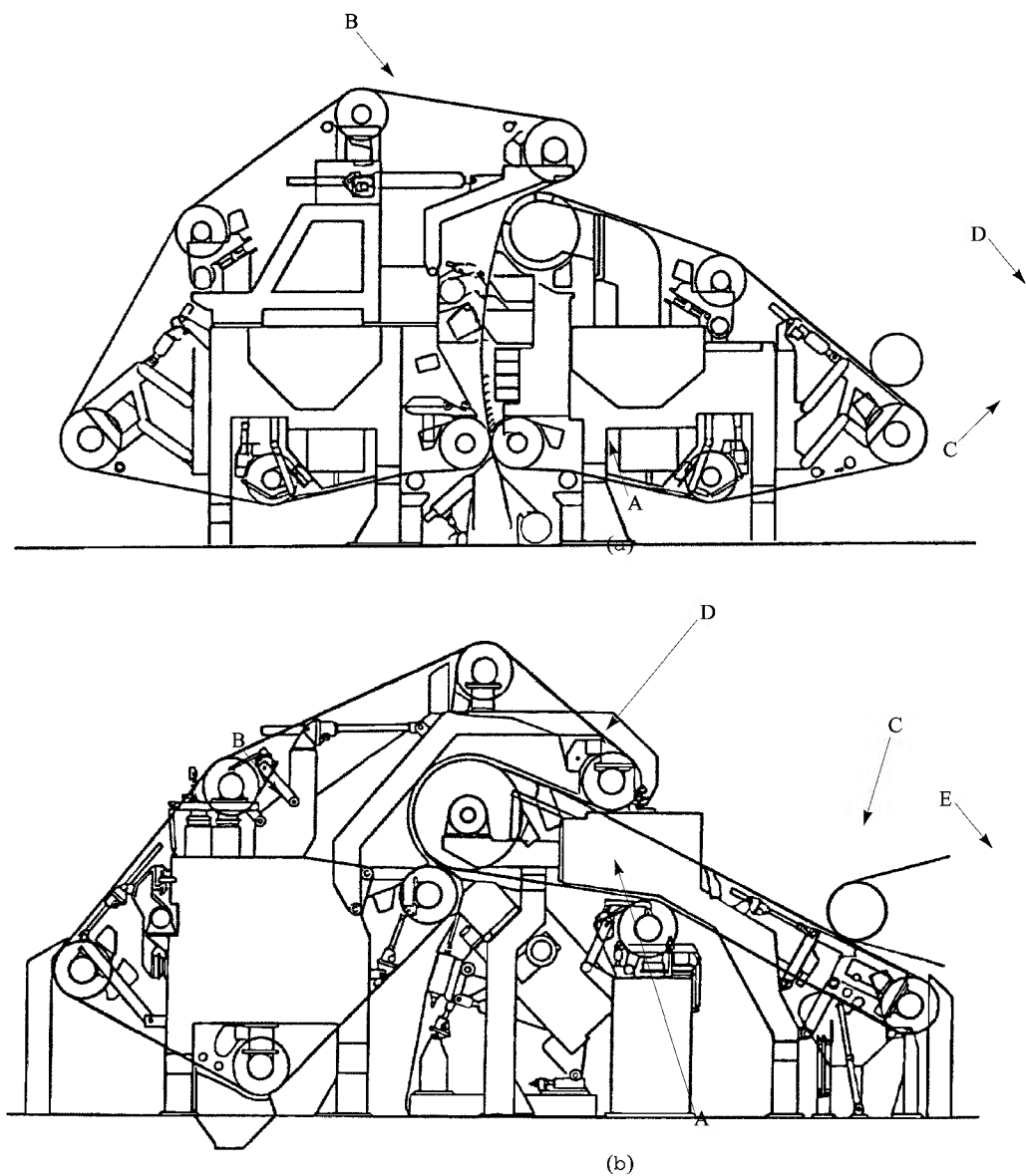


Fig. 6. (a) The Bel Baie II twin-wire former. Stock from A, the headbox, is formed into sheet between B, wire number 1 and C, wire number 2. The web is removed from wire 2 by D, a suction pickup roll. (b) A roll-type twin-wire tissue former. Stock from A, the headbox, passes between B, wire number 1 and C, wire number 2 around D, a forming roll, and the web is removed from wire 2 by E, a suction pickup roll.

Courtesy of Beloit Corp.

Boxboard normally consists of 3–7 separate webs that are formed from different raw materials and couched together. Because of the number of plies and wide variations in speed requirements, there are a large number of different designs for board-forming sections. The classical forming section for a board machine is the vat machine (Fig. 7). The incoming diluted slurry is introduced to the cylinder vat after final screening. The sheet is couched onto the underside of a long carrier felt or onto a previously formed paper layer which is carried by the felt. Because of the limited drainage capacities of the vats and the method of carrying the sheet, vat machines are limited to low speeds. Modern versions of the vat machine employ cast and drilled shells, suction boxes, and pressurized headboxes. These additions permit increase in the drainage capacity and facilitate water handling, so that the formers can be placed on top of the carrying fabric felt for higher speed operation.

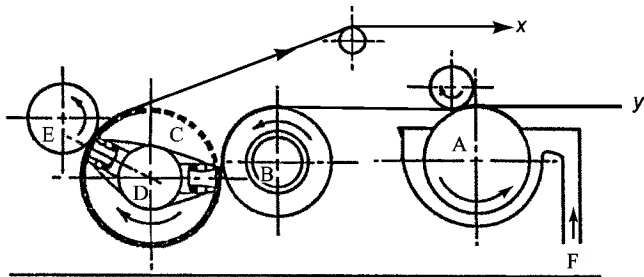


Fig. 7. The vat machine. A, last cylinder taking up slurry from the surrounding vat; B, extractor; C, perforated cylinder containing D, a stationary water extractor; E, press; F, supply of slurry; x, paper on its way to the dryers; y, multiple wet paper sheet coming from preceding cylinders (15).

Courtesy of J. H. de Bussy, Amsterdam, the Netherlands.

Another development is the Inverform process and its more modern version, the Bel Bond (Fig. 8). In the Inverform unit, several plies are formed on top of each other by consecutive, twin-wire forming units above a long carrying fabric. The Inverform process also is used for the forming of paper grades and is capable of moderately high speeds. Other versions of board machines involve mini-Fourdriniers and/or twin wires which are placed on top of a carrying fabric.

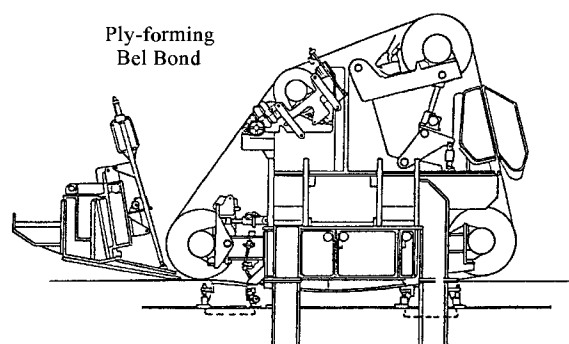


Fig. 8. One Bel Bond unit of a multi-ply board machine.

Courtesy of Beloit Corp.

All the preceding forming units receive the incoming slurry at a low consistency, typically 100–300 kg water/kg solids, and the paper web leaving the couch typically contains 4 kg water/kg solids. The white (drained) water contains some fiber debris, clay filler, etc, and is reused for dilution of the incoming stock. In closed paper machine systems, any excess white water is filtered and the recovered solids are returned into the system. In order to reduce the amount of water that must be drained from the sheet, a development called a high consistency former is sometimes employed. By virtue of narrow forming channels, high consistency headboxes can generate a scale of turbulence that is fine enough to break up flocs of fibers in stock at 3–4% solids, thereby reducing the amount of water that has to be removed by a factor of 10. Because the forming gap is proportional to the weight of the ply being formed, the channel width in a high consistency former becomes too fine for practical operation at ply or sheet weights below about 60 g/m². Paper made with high consistency formers has a characteristic oatmeal formation since all of the flocs are of the same dimensions as the width of the forming channel.

Sheet Pressing. The sheet leaving the wet end contains on the order of five parts of water per part of fiber; however, it is possible to remove additional water mechanically without adversely affecting sheet properties. This is achieved in rotary presses, of which there may be one or more on a given paper machine. The press rolls may be solid or perforated and often suction is applied through the interior. The sheet is passed through the presses on continuous felts (one for each press), which act as conveyors and porous receptors of water. They are essential to the efficiency of the papermaking process since it is much more expensive to remove water from paper by drying with heat than by dewatering with pressure. The water content of the sheet usually can be reduced by pressing to 1.9–1.2 parts of water per part of fiber without deleteriously affecting product quality.

Novel press configurations which extend the time under pressure can achieve higher solids ratios. It is possible to combine pressure and drying in a press where one of the rolls is heated well above the boiling point of water, usually referred to as impulse drying. Hot steel belts can also be used to dry the paper under pressure. In both of these press drying configurations the paper tends to be more dense than with conventional pressing and drying since the fiber network is being compressed as it is being softened at high temperatures. Thus, initial applications of press drying have been to products where increased sheet compaction and bonding can be of value, such as linerboard. There is evidence to suggest that at sufficiently high temperatures, lignin in the high yield kraft pulps used for linerboard may flow somewhat to impart properties such as greater stiffness and strength when wet.

Sheet Drying. At a water content of ca 1.2–1.9 parts of water per part of fiber, additional water removal by mechanical means is not feasible and evaporative drying must be employed. This is at best an efficient but costly process and often is the production bottleneck of papermaking. The dryer section most commonly consists of a series of steam-heated cylinders. Alternate sides of the wet paper are exposed to the hot surface as the sheet passes from cylinder to cylinder. In most cases, except for heavy board, the sheet is held closely against the surface of the dryers by fabrics of carefully controlled permeability to steam and air. Heat is transferred from the hot cylinder to the wet sheet, and water evaporates. The water vapor is removed by way of elaborate air systems. Most dryer sections are covered with hoods for collection and handling of the air, and heat recovery is practiced in cold climates. The final moisture content of the dry sheet usually is 4–10 wt %.

Other types of dryers may be employed for special products or situations. For example, the Yankee dryer, a steam-heated cylinder, 3.7–6.1 m dia, dries the sheet from one side only. It is used extensively for tissues, particularly where creping is accomplished as the sheet leaves the dryer, and to produce machine-glazed papers where intimate contact with the polished dryer surface produces a high gloss finish on the contact side.

High velocity air drying, in which jets of hot air are directed against the sheet in a normal direction, is used in Yankee dryers and in combination with a percolation through-drying process. In the latter, hot air is drawn through the sheet, thereby effecting a high heat transfer to the sheet and efficient mass transfer of the water vapor from the sheet. The latter technology is commonly used for special, high quality tissue products. Infrared and other radiant drying techniques also are utilized in special cases.

Conventional cylinder drying sections employ two felts to hold the paper web against the drying cylinders, one on the bottom tier of cylinders and one on the top tier. In this configuration the sheet is free to shrink laterally between dryer cylinders, and thus the paper web is restrained less than 40% of the time. This, together with the tendency of fibers to align in the direction of forming, leads to significant differences in physical properties of machine-made paper in the manufacture direction (MD) and in the cross direction (CD). For example, modulus of elasticity and tensile strength are typically more than twice as high in the MD. Hygrostability of paper also depends strongly on the degree of restraint during drying, therefore it is of interest to design drying sections which hold the paper under CD restraint as far as possible. Single-tier dryer sections hold the paper under CD restraint about 50–60% of the time and reduce instabilities such as web flutter (16). It is also possible to introduce suction rolls between the drying cylinders and achieve about 80% restraint. Ultimately, the greatest restraint during drying occurs in Yankee and machine glaze (MG) drying cylinders where the sheet adheres to the dryer until dry. Papers such as glassine made on MG machines are therefore more dimensionally stable under conditions of changing relative humidity than fine paper made on conventional dryers.

Converting. Almost all paper is converted by undergoing further treatment after manufacture. Among the many converting operations are embossing, impregnating, saturating, laminating, and the forming of special shapes and sizes, eg, bags and boxes (see PACKAGING, CONVERTING).

Pigment Coatings. Pigment coatings are compositions of pigments and binders with small amounts of additives and are applied to one or both sides of a paper sheet. These generally are designed to mask or change the appearance of the base stock, improve opacity, impart a smooth and receptive surface for printing, or provide special properties for particular purposes. Pigment coatings are highly porous because the binder adhesive is insufficient to fill the void spaces among the pigment particles. This porous structure is responsible for many of the desired properties of the coated papers. For example, the high opacity of the coatings results from the scattering of light from the pigment–air and binder–air interfaces (see also COATINGS; COATING PROCESSES; PIGMENTS). In many instances, titanium dioxide is used to enhance opacity. Because of the high refractive index, TiO₂ increases opacity and light can be scattered at the interface of TiO₂–air as well as at the interface of lower refractive index materials such as binders and other pigments.

Pigment-coated printing papers usually are required to have high brightness to achieve contrast between the printed and unprinted areas. The coatings frequently have a glossy surface. There also is a demand for dull-coated papers and for dull papers upon which glossy ink films can be printed. The coated paper should be sufficiently smooth for printing to allow full contact between the inked image area of the plate or transfer blanket and the paper surface. Requirements for smoothness decrease from gravure printing to polymer plate flexography to offset lithography. Printing smoothness involves not only the smoothness and compressibility of the paper, but the amount of ink, the properties of the ink-transfer and impression materials, the printing pressure, and the printing process. Consequently, smoothness is best predicted by tests that simulate the printing conditions for which the paper is intended (TAPPI UM466 and UM505). Because only the coating layer is involved significantly in absorption of the small amount of ink fluid applied, tests

that involve transudation of the sheet cannot be expected to provide pertinent information on ink absorbency. Comparison of ink absorbency of papers can be made according to TAPPI UM553.

The printing process imposes tensile stress normal to the plane of the sheet; the stress depends on the tack of the printing ink and the velocity of separation of the printing plate from the paper. The stress also tends to pick the paper, ie, remove material, unless the paper has adequate pick strength. Papers to be printed by offset lithography require high pick strength because of the tacky inks employed, whereas gravure papers need not be so demanding. In both cases, papers to be used for multicolor printing require higher pick strengths than those used for single-color printing because of the range of tack required in the inks (qv). For example, gravure papers do not require high surface strength because of the low viscosity inks used (see PRINTING PROCESSES). To an increasing extent, pick strength is determined using printing tests that employ tack-graded inks or viscous test liquids at a controlled printing speed and pressure (see TAPPI T499 and UM507). The pick strength of a coating depends on the pigment and the type and amount of adhesive used as binders. The desired pick strength usually dictates a minimum amount of binder.

In offset printing, some water is transferred from the blanket to the paper. If the binder is water soluble, the dampened coating is more susceptible to picking at subsequent impressions. Consequently, offset papers for multicolor printing must have water-resistant coatings. The water resistance is achieved by using insolubilizing agents in the adhesive (see TAPPI UM513).

Application. Pigment coatings normally are applied to the base paper in the form of water suspensions and are referred to as coating colors. The total solids content, ie, pigment plus binder, may vary widely. Blade coatings range from 50 to 70% ; pigmented size coatings vary from 6 to 15% . After application, the coating must be dried by removal of water from the film. In some cases, a calendering operation serves to smooth the surface, control surface texture, and develop a glossy finish.

Paper may be coated either on equipment that is an integral part of the paper machine, ie, on-machine coating, or on separate converting equipment. Many plants include both types of coating equipment and utilize each to its maximum advantage for paper and paperboard. The combination of techniques is of particular value where more than one coating must be applied to the sheet in order to obtain a product of desired quality.

In 1933, the first roll coater was installed as an integral part of a paper machine. These on-machine coaters produce a low cost coated paper used largely for magazines. Coating of paper off-machine at speeds greater than paper machine manufacturing speeds is possible with the use of the various available blade coaters (17).

There are numerous roll coaters by which coatings are applied from rolls which travel at paper speed. Conventional size presses, in which paper passes through open puddles of aqueous solutions of starch with and without hydrophobic additives, are giving way to metering size presses where starch and hydrophobes are metered to a thin film on the size press cylinders by blades or rods and then transferred to the paper. It is also possible with modern metering size presses to apply pigmented coatings. Metering size presses can thus provide a form of inexpensive roll coating on-machine. However, patterns can develop in the coated surface because of the way the film of coating color splits between the roll and paper surfaces. Therefore, it is important to formulate these roll coatings so that leveling of the coating surface occurs before drying takes place.

Using air doctor coaters, a coating color is applied with a roll and the excess is removed with an air doctor, ie, a long thin jet of air which acts as a doctor blade (Fig. 9a). A coating of uniform thickness is achieved and surface contours tend to follow those of the raw stock. The coating colors usually are more fluid than those used in other coating methods. With blade coaters (Fig. 9b), the coating is smoothed and excess coating color is removed by a flexible blade. The blade is supported by the paper which tends to fill the depressions with a troweling action, thus much less coating color is applied to the high points of the sheet. It is useful where the levelness of the coated surface is more important than uniform thickness of coating. Coating colors of high solid content are used with blade coaters. Because less water must be removed, the coatings can be dried at high speed.

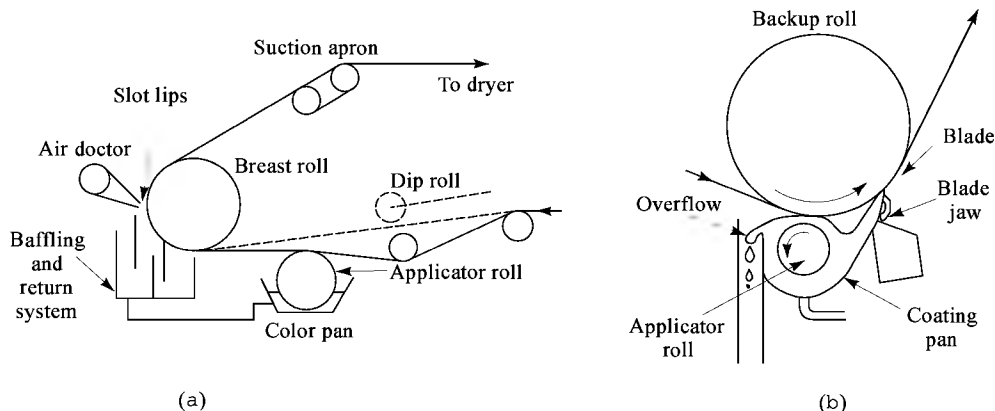


Fig. 9. (a) Air-knife coater; (b) inverted blade coater.

In cast coating, the coating color is pressed in contact with a highly polished metal drum immediately after application. The coating is dried in contact with the drum to give a smooth glossy surface, and subsequent calendering is unnecessary. All pigment coatings must be dried to remove the water from the coating and the water that has penetrated the sheet. Drying methods include air or convection drying, contact or conduction drying, and radiant-energy drying. The speed of a coating operation often is restricted by the rate at which the water can be removed from the coating without blister formation or excessive migration of components to the hot surfaces. Coated paper also is supercalendered to improve surface properties, eg, smoothness and gloss. The paper is passed successively through the nips of a stack of alternating hard steel and soft rolls. The pressure causes the soft roll to deform, which burnishes the surface of the sheet against the polished steel surface. This action softens the coating binder and causes greater alignment of clay platelets and produces a glossy level surface. The properties of the final surface are functions of the coating composition, number of rolls, pressure, temperature, and operation speed. The process also tends to densify the sheet and thus to reduce porosity (see COATING PROCESSES).

Pigments. Pigments comprise 70–90% of the dry solids in paper coatings (18). In nearly all cases, the individual particles of the pigments are less than 5 μm in equivalent spherical diameter and average less than 1 μm. Many pigments are less than 0.5 μm in size. These particles can fill the spaces between fibers on the sheet surface and form a nearly uniform surface mat. Pigments also control opacity, gloss, and the color of the raw stock. Refractive index, particle size, crystal structure, light scattering and absorption, and adhesive demand are important characteristics. Minimum amounts of binders are used to bind the pigments, because the binders are more expensive than the pigments and excessive amounts can adversely affect the opacity and brightness of the coatings by filling in the air–pigment interfaces that scatter light. Kaolin clays are hydrated aluminosilicates having the generalized formula $(Al_2(OH)_4(Si_2O_5))_2$. They occur as natural hexagonal plates with a diameter-to-thickness ratio of ca 10:1. Variations in the specific makeup of the clays influence rheological, surface, chemical, and optical properties. Minor amounts of impurity minerals can adversely affect pigment properties (see CLAYS).

Commercial coating clays have been fractionated by centrifugation to remove grit and large particles and to obtain an optimum range of particle sizes. Special treatments improve the brightness and gloss properties and adjust the viscosity of the clay–water suspensions. Coating clays are produced in several grades, depending on the properties desired. Brightness may vary from 75–85% for filler grades to greater than 90% for special coating grades. The particles average ca 0.8 μm in equivalent spherical diameter, but because of the plate form, the kaolins produce glossy surfaces after calendering of the paper. The pigment produces many surfaces which scatter light and thus it contributes to opacity even though its refractive index is only 1.55, which is about the same as the cellulose fibers it must cover. Clays are the most common and most widely used pigments employed in paper coating.

The average particle size of coating-grade titanium dioxide is ca 0.3 μm. Because this size is optimum for maximum hiding power and because of its

high refractive index, titanium dioxide pigment has a unique capacity to opacify and brighten coated paper. The pigment has a brightness greater than 95° o. It is chemically inert and is easily dispersed in water. Both anatase and rutile crystal forms are used. The rutile is preferred in situations where waxed or saturated opacity is important because of a slightly higher refractive index. Partly because of high cost, the amount of titanium dioxide in a coating seldom exceeds 25 wt % of the total pigment. In high quality products where uv or fluorescent brighteners are used, rutile, a particularly strong uv absorber, adversely affects the brightening.

Calcium carbonate, available both from natural sources and as precipitated forms (see [CALCIUM COMPOUNDS](#)), is most useful in coating because of purity and high brightness, ie, 90–95° o. Ground carbonates from marble deposits have high purity levels as do the carbonates from some chalk deposits. By changing the precipitation conditions, particle size and shape can be varied over a wide range, and recipitated calcium carbonates can be optimized to scatter light, provide an appropriate pore system, or to provide bulk and a low abrasive value. When used with clays to produce improved brightness carbonates disrupt the orientation of the clay plates, reduce the surface gloss, and maintain the porosity of the coating. Carbonates, whether precipitated or from natural sources, are a common component of matte-finish coatings. In Europe, calcium carbonate is used even more extensively than kaolin in high quality coatings.

In many coating formulations, a combination of binders is used to maximize the properties of various types of binder systems. The amount of binder in a coating formulation varies from about 5 to 25° o, based on the pigment. The actual amounts depend on properties desired in the coating, specific use of the sheet, and types of binders to be used (see [ADHESIVES](#)). Animal glue was the first material used for bonding paper and as an adhesive in paper coating. The use of glue is confined to specialty applications in paper converting.

In the late 1800s, when the demand for coated paper for the halftone printing process increased, casein rapidly replaced glue. Casein forms a hard, tough film when dry, and can be waterproofed easily with formaldehyde (qv). The properties of soy protein are similar to those of casein, and soy protein has been substituted for it in many types of coated papers requiring a casein-type binder (see [SOYBEANS AND OTHER OILSEEDS](#)). Casein, a valuable food product, is seldom used as a paper adhesive, in spite of its excellent adhesive properties.

Starch (qv), obtained from corn, potatoes, tapioca, wheat, etc, is a polymer of glucose in which the units are arranged in the linear amylose and the branched amylopectin forms. Because of the complexity and size of the structure and the high viscosity of starch in solution, starches cannot be used directly in coating formulations, but must be modified to produce a lower viscosity system. Modification of starch by enzymes or thermal conversion is common. Acid-hydrolyzed starches that have been hydroxy ethylated or acetylated are also common, as are hypochlorite-oxidized starches. Hypochlorite-oxidized starches are characterized by lower gelatination temperatures, clearer solutions, and lower viscosities at high concentration than their parent forms. Oxidized forms are available in a range of viscosity grades and are used in sizing operations as well as in pigment coating.

Starches are quite hydrophilic and the granules swell in water to several times their volume. In the preparation of starches for coating, the granules are heated at 93°C to ensure complete breakdown. The starch is then mixed with the pigment while it is fluid. Normal practice is to coat the paper at an elevated temperature to help control viscosity.

Starches can be made more resistant to water by the use of aldehyde-donor cross-linking agents, eg, urea–formaldehyde and melamine–formaldehyde or glyoxal; however, these materials increase viscosity. Exceptionally strong binders, eg, poly(vinyl alcohol) (PVA), can be used in coating formulations. The strength of PVA in a coating is approximately four times that of starch and three times that of casein. Although PVA is expensive, only small amounts are required and it is characterized by good optical properties. However, its high solution viscosity limits its use in coating formulations.

Various rubber latexes and other emulsions are commonly used as binders in high speed coating operations. Such materials can be added to the coating without special preparation, as is required with the natural adhesives. Emulsions provide many properties that are superior to those of the aforementioned natural binders, eg, low viscosity which permits high solids content, easy handling, and low water content. Emulsions and latexes are distinguished by high gloss, good response to calendering, good ink holdout (nonblotting character), and good water resistance. The styrene–butadiene latex is commonly supplied in a 60:40 ratio of styrene to butadiene. It is used with starch and primarily in publication-grade papers (see [ELASTOMERS, SYNTHETIC; LATEX TECHNOLOGY](#)).

In the United States, acrylic-based emulsions are used mostly on paperboard. These materials are odorless, which is necessary for coatings to be used on paperboard for food packaging (qv). The acrylics provide high gloss and good ink retention. However, use of these materials has been restricted because of high cost. Poly(vinyl acetate), the adhesive strength of which is equivalent to the previously mentioned emulsions, is used in place of acrylics. Poly(vinyl acetate) also provides moisture and grease resistance. In Europe, acrylic is commonly used in paper coatings.

Additives. Additives control coating behavior during application or they can be used to alter the properties of the finished product. A single chemical additive may be used for several purposes. Some additives are essential to the production of a salable product, and others may be added only to obviate problems of the coating operation.

A dispersing agent is used to transform pigments into a slurry form. The material, usually a polyphosphate, is adsorbed on the pigment and causes the particles to repel one another, thereby reducing the coating viscosity. Proteins and casein also must be dispersed. The various emulsions contain stabilizers and compatibility between these stabilizers and other coating components must be considered for all coating systems (see [DISPERSANTS; EMULSIONS](#)).

Foam-control agents commonly are employed (see [DEFOAMERS](#)).

Lubricants, plasticizers (qv), and flow modifiers include both soluble and insoluble soaps, sulfated oils, wax emulsions, amine products, esters, etc. Certain materials, eg, urea and dicyandiamide, may reduce the viscosity of a coating color (see [CYANAMIDES](#)). Lubricants improve flow, coating smoothness, finish, printability, and antidusting effects (see [LUBRICATION AND LUBRICANTS](#)). Humectants, eg, glycerol derivatives, are used in small amounts as plasticizers and aid in the development of the finish on the sheet. Materials used to increase the moisture resistance of a coating surface or to insolubilize the adhesive may not have film-forming properties. They may react with the hydrophilic groups in the adhesive or cross-link the polymer chains to prevent swelling with water and subsequent loss of binding strength. The decrease in water sensitivity can be achieved by mixing formaldehyde donors, eg, urea– or melamine–formaldehyde resins (see [AMINO RESINS AND PLASTICS](#)), or glyoxal with the coating formulation. In some cases, the desired results may be obtained by exposing the surface to formaldehyde or by application of a zinc, aluminum, or other metal–salt solution during calendering. Some resistance to moisture always is obtained when a latex is included in the coating formulation.

Barrier Coatings. In packaging applications, a barrier may be needed against water, water vapor, oxygen, carbon dioxide, hydrogen sulfide, greases, fats and oils, odors, or miscellaneous chemicals. A water barrier can be formed by changing the wettability of the paper surface with sizing agents. A grease or oil barrier can be provided by densifying paper made from well beaten fiber to form a pinhole-free sheet or by coating the paper with a continuous film of a material resistant to the particular grease. Gas or vapor barriers are formed by coating the paper with a continuous film of a suitable material (see [BARRIER POLYMERS](#)).

Paraffin wax is applied in a molten form. It is low in cost, resists water vapor and is colorless, and is free from odor, taste, or toxicity. It is applied by passing the paper through a molten bath or nip, removing the excess paraffin, and chilling. Modifiers, eg, microcrystalline wax, polyethylene, or ethylene–vinyl acetate copolymer, improve the durability and film strength, raise the softening point, and increase the gloss and heat-seal strength of the coating. Polyethylene coatings are more durable and flexible, and thus have largely replaced wax coatings. Polyethylene and other barrier polymers such as ethylene–vinyl alcohol are applied by extrusion or coextrusion with intervening tie layers. Polymer pellets are heated rapidly with minimum air contact, and each molten polymer is extruded through a die and immediately laminated to the paper.

Solvent systems permit the formulation of highly sophisticated coatings, comprised of a wide variety of polymers and various modifiers. Disadvantages include high solvent costs and the necessity for a solvent recovery system. The resins used include many of the same types used for coated fabrics or for industrial coatings, eg, cellulose derivatives, rubber derivatives, butadiene–styrene copolymers, vinyl copolymers, poly(vinylidene chloride), polyamides, polyesters, and alkyds. High solids content at minimum viscosity is available with emulsion or latex coatings. Poly(vinylidene chloride) provides excellent barrier properties (see [VINYL POLYMERS; VINYLIDENE CHLORIDE AND POLY\(VINYL CHLORIDE\)](#)).

Economic Aspects

As of 1995, the forest industry in the United States employed about 1.6 million people and produced products valued at over \$200 billion each year, approximately \$20 billion of which was in exports. There were 350 pulp mills, 600 paper and board mills, and ca 4500 converting plants in the United States producing ca 30% of the total world production. The United States and Western Europe, which represent ca 13% of the world population, consumed about 60% of production.

The total production of the U.S. paper industry in 1994 was ca 85 million metric tons with a compound annual growth rate over the previous decade of 2.7%. The domestic production of paper and board plus imports and minus exports maintains a remarkably constant ratio with real gross domestic product in the United States. One sector of the paper industry that has grown at a higher rate than GDP is recycled papers and boards which is projected to grow at 6.8% annually. Over one-half of paperboard production in the United States is from recycled fiber, and the industry as a whole is expected to achieve 50% recovery rate for paper and board products by the end of the twentieth century (see RECYCLING, PAPER).

Printing and writing papers continue to gain share at the expense of packaging and industrial papers. This trend is likely to continue as digital printing creates markets including international digital distribution of information and local printing, on-demand printing of books and manuals, and customization of magazines, catalogues, and advertising. Newsprint has maintained its share of U.S. production mainly through a reduction in the amount of newsprint imported from Canada, which has declined to less than half of the U.S. newsprint consumption. Economic and statistical information on the U.S. pulp, paper, paperboard, and allied industries is given in Reference 19.

Analytical Test Methods

Chemical Composition. Methods for paper analysis are reviewed and experimental details are provided in Reference 20. Chemical analyses which are used to characterize wood pulps for papermaking include determinations of α -, β -, and γ -cellulose, carbohydrates, lignin, carboxyl and carbonyl groups, copper number, and viscosity. The carbohydrate determination involves acid hydrolysis of the pulp, preparation of volatile derivatives, and separation of the individual monomeric sugars by gas chromatography (see TAPPI T249 and ASTM D1915). Results are used to compute percentages of cellulose and hemicellulose in the pulp. Lignin is that portion of the pulp that is insoluble in 72 wt % sulfuric acid (see TAPPI T223 and ASTM D1106). The α -, β -, and γ -cellulose test (TAPPI T203) and other determinations based on alkali solubility (TAPPI T212 and T235, ASTM D1696) reflect empirically the hemicellulose content of the sample and the degradation of the cellulose. Also indicative of hydrolytic or oxidative degradation of the pulp are higher carboxyl (TAPPI T237 and ASTM D1926), carbonyl, and copper number values and lower viscosity. Cupriethylenediamine solutions of cellulose are used for the viscosity test (TAPPI T254 and ASTM D539); results are related to the average degree of polymerization. Molecular weight distribution may be determined by separation of cellulose tricarbanilate derivatives by gel-permeation chromatography (gpc) (21).

Procedures are available for detecting and determining most of the noncellulosic constituents of papers. Rosin size is detected by the Raspail or Lieberman-Storch test; it may be determined by extracting the paper with acidified alcohol and isolating the ether-soluble portion of the alcohol extract (TAPPI T408 and ASTM D549). Starch is detected by the blue color produced with application of an iodine–potassium iodide solution. The intensity of the blue, which is measured with a spectrophotometer, provides the basis for a quantitative starch determination (TAPPI T419 and ASTM D591).

Kjeldahl nitrogen determinations are used to determine the wet-strength resins most commonly used in paper, eg, urea–formaldehyde and melamine–formaldehyde (TAPPI T418 and ASTM D982). Melamine may be determined by uv spectrophotometry (TAPPI T493 and ASTM D1597). Formaldehyde-containing wet-strength resins can be detected if a red-violet color appears after heating the paper in a solution of chromotropic and sulfuric acids (see AMINO RESINS AND PLASTICS).

Acidity or alkalinity of paper is determined by measuring the pH of a cold- or hot-water extract (TAPPI T435 and T509, ASTM D778). Alum is the most common source of acidity. The amount of mineral filler or coating pigment is determined from the ash content of the paper (TAPPI T413 and ASTM D586). Factors may be necessary to correct for ash in the pulp and for pigment changes resulting from ashing. Distribution of filler through the thickness of the web is measured by removing increasingly greater amounts of paper by surface grinding and determining ash in the remaining paper. Pigments usually can be identified by x-ray diffraction analysis. Elemental analysis by emission spectrography or energy-dispersive x-ray analysis can aid in pigment identification.

Analysis of coatings is simplified if the coating can be removed from the paper in a water bath by ultrasonic cleaning. If fibers are not present, a carbohydrate determination can be used to identify gums and other carbohydrate polymers in the coating.

Latexes of synthetic resins are identified by ir spectrometry. Selective extraction with organic solvents is used to obtain purified fractions of the polymers for spectrometric identification. Polymeric films can be identified by the multiple internal reflectance ir technique, if the film is smooth enough to permit intimate contact with the reflectance plate. TAPPI and ASTM procedures have not been written for these instrumental methods, because the interpretation of spectra is not amenable to standardization.

Solvent extraction followed by gas chromatographic analysis is used to determine paraffin wax; antioxidants (qv), ie, butylated hydroxyanisole and butylated hydroxytoluene; and other volatile materials. Trace amounts of chlorinated organic compounds, eg, polychlorinated biphenyls, can be determined by using a gas chromatograph with an electron-capture detector (22).

Fiber Analysis. Paper may be composed of one or several types of fibers, eg, animal, vegetable, mineral, and synthetic (see FIBERS). Paper is generally composed of woody vegetable fibers obtained from coniferous (softwood) and deciduous (hardwood) trees. Qualitative and quantitative methods have been developed to determine the fibrous constituents in a sheet of paper (see TAPPI T401). However, the proliferation in the number and types of pulping processes used have made the analysis of paper a much more complex problem. Comprehensive reviews of the methods are given in References 20 and 23.

A common method involves tearing a representative sample into small pieces, placing these in a beaker, and heating to a boil in 0.5% sodium hydroxide. The paper then is washed, neutralized with hydrochloric acid, and washed again. Disintegration is effected by vigorously shaking the flask. A desirable fiber concentration for suspension is ca 0.05%. A portion of this suspension is placed on a slide and examined microscopically. The fibers are stained in order to determine the pulping process and to produce contrast for the identification of the fibers. Graff's C stain is probably the best stain for general fiber analysis. Other stains used are Wilson's stain, Herzberg stain, and the Green-Yorston stain, AZO. The colors that are developed by these stains vary according to the raw material used and the pulping process. The fibers should be examined in daylight or fluorescent lighting at a magnification of ca 100 diameters.

Analysis of certain papers requires special treatment before they can be disintegrated properly. Papers containing synthetics, tars, asphalt, rubber, viscose, or wet-strength resins must be analyzed individually (see TAPPI T401) (20). Dyes or colors must be removed from highly colored papers before examination. The method of dye removal depends on the type of dye.

Plant-fiber identification is described in TAPPI T8 and T10. In order to identify synthetic fibers, it usually is necessary to conduct solubility and physical properties tests in addition to light microscopy observations. Systematic sampling is required to obtain quantitative information on sample composition. Because different types of pulps contain varying numbers of fibers per unit weight, it is necessary to multiply the total number of each kind of fiber by a relative weight factor, thereby the weight percentage that each fiber type contributes to the sample can be determined.

Environmental Issues and Plant Efficiency

Modern practice is to maintain the white water system as closed as possible, ie, as much water as is compatible with efficient machine operation is recycled. The loss of fibers and inert furnish components, particularly clay, has been greatly reduced. Fiber losses, however, still occur into the white water, and greater economy of operation may be achieved if these fibers could be recovered. Thus, it is common to design a fiber-recovery system into the white water cycle. The three general types of save-all fiber recovery are based on filtration (qv), flotation (qv), and sedimentation (qv). If these are operated efficiently, the net fiber loss can be less than 1%.

In order to conform to environmental quality guidelines, mills have installed a number of primary and secondary treatment systems to control

effluents (24). The primary treatment is composed of settling basins and or tanks, ie, clarifiers. These remove ca 85–100 wt % of solids, eg, fibers and clay. Primary sludges, which are removed by the primary clarifiers, cannot be reused by the mill in the same product (25). However, many mills use the sludges in lower grade product lines. The secondary treatment generally consists of a biological treatment followed by secondary clarifying. Biological waste treatments include lagoons, aerated lagoons, activated sludge with air and oxygen, trickling filters, modified biological systems containing activated carbon, and combinations of these. The secondary treatments generally remove 90–95% BOD, most solids, most of the toxicity, but little color. In some instances, color increases after the water has been treated. Some of the solids from the secondary clarifiers may be recycled but some is wasted. These solids or biological sludges are extremely hydrophilic and are difficult to dewater.

As more process water is recycled to reduce overall water consumption and wastewater discharge volumes, more nonpathogenic microbial growth, ie, slime, occurs in the mill system. Slime formation prevents normal flow of stock suspensions, may make the furnish lumpy, prevents normal sheet formation, and interferes, in general, with papermaking. It has been a significant impediment to the goal of 100% closure of the water loop. Fundamentally, the remedy for slime problems is to create conditions in the system that are necessary for the growth and propagation of slime-forming organisms. The efficiency of any form of slime control is greatly increased by ordinary good mill cleaning procedures. The application of chlorine or chloramine with or without frequent cleaning is effective in many cases. Antiseptics and disinfectants reduce or inhibit slime formation (see DISINFECTANTS AND ANTISEPTICS). The use of these slimicides may result in an appreciable increase in the cost of producing paper. However, their use often reduces downtime that is caused by slime and, therefore, increases production which more than compensates for the initial cost of the slimicides. White water systems also often contain proteolytic microorganisms which attack the machine felts and reduce their useful life. Control of this problem may be accomplished by treating the felts with a slimicide followed by cleaning with a mild acid (see INDUSTRIAL ANTIMICROBIAL AGENTS).

Sludge Handling and Disposal. Most waste-treatment processes generate solid wastes that must be disposed (26). Two kinds of sludges are generated by pulp and paper mills: primary sludges contain fibers, clay filler materials, and other chemical additives; secondary sludges are largely biological in nature and harder to handle and dewater (27). The disposal of sludges in landfills is being reevaluated and alternative disposal approaches are being developed (28–30).

The wide range of types of paper products results in a variety of sludges. Solid wastes result from several sources within the mill, eg, bark, sawdust, dirt, knots, pulpwood rejects, flyash, cinders, slag, and sludges. Sludges often are disposed of in combination with residuals from other sources. Approximately 300 kg of solid waste per ton of finished product is generated by the pulp and paper industry.

Solids content of wet sludges is 1–40 wt %. Ash content can vary from very little to over 50 wt % of the solids content. However, the solids or moisture content of a sludge is not enough for assessing the physical or engineering properties of that sludge. For example, sludges of 30–35 wt % solids from a paper recycling or deinking operation may be in a highly fluid state, whereas a low ash, high fiber pulp mill sludge at 15–20 wt % solids may be quite dry and stable. The operating plan for a land-disposal site depends on the fluid state of the sludge (29). Generally, sludge handling processes include thickening, stabilization, conditioning, dewatering, incineration, and disposal (31). Most sludges are disposed in landfills.

Water Quality Assessment. Assessments of the effects of effluents on receiving streams until the mid-1970s were more subjective than objective (32,33). Since then, changes in attitude within the aquatic life-science field and the regulatory system have necessitated a restructuring of the design and foundations for effluent-impact assessments in receiving waters. For example, government regulations have proceeded from a system of stream standards-based regulations to effluent-based regulations involving strict requirements on various pollution parameters. Pressure is being exerted to go back to the receiving water system as the ultimate test for new and more stringent discharge-control measures. The increasing knowledge of the interrelationships between the various biological, chemical, and physical components of aquatic systems has provided significant restructuring of field assessment programs which are designed to analyze effluent impact. A single organism as an indicator of stream quality has been replaced by community compositional and structural analysis. Thus, total effluent effects on a broad scale can be realized. Other measured parameters are algal assays, fish surveys, sediment mapping, plume mapping, sediment oxygen demand, and socioeconomic impacts.

Uses

Paper and Paperboard Containers.

Rigid Paper Containers. Rigid paper containers generally are constructed of paperboard or a combination of paperboard and paper.

Setup Boxes. The principal requirement of the setup box is stiffness. Because setup boxes are used for candy, stationery, etc, they need not be very strong. They constitute only a small percentage of total box production. Setup boxes usually are formed from a blank of single-ply, stiff paperboard, although pasted boards sometimes are used, by cutting the board from the outside almost entirely through its thickness along lines which are intended to be the edges of the final box. The board is folded along the precut lines, and the edges which are formed by the cut score, as well as those which are formed where the sides of the cut blank meet, normally are taped with paper to hold the box together and to reinforce the cut edges. A cover paper usually is glued to the outside surface of the box. Setup cartons are assembled in the manufacturer's plant before shipping.

Folding Cartons. Folding cartons differ from setup boxes in the type of paperboard used and the method of creasing. Because of its high stiffness, the paperboard is creased by means of a scoring or creasing rule. The board is crushed in the area of the crease and subsequently is folded along these predetermined crushed lines. Folding cartons are shipped flat and must be opened and set up for use.

The type of paperboard used by the carton industry is boxboard. Boxboard may be categorized, based on the raw material, as combination or solid boxboard. Combination boxboard, of which there are many grades, normally is made on a multicylinder paper machine using a substantial percentage of waste paper with virgin pulp. Solid boxboard usually is made on a Fourdrinier paper machine using only virgin pulp and it is bleached or coated.

Although folding cartons are made in many sizes and shapes, all are of three basic designs. Tray cartons are open on one face and are formed by folding a sheet of board to make the side panels. Covers are glued or locked in place. Top-opening cartons are similar to tray cartons, except that one side panel is extended to serve as the top. This is folded over to cover the open face. The cartons may be tucked, locked, or glued closed. End-opening cartons are essentially tubes in which one or both ends are folded and sealed, locked, or tucked closed. The folding carton primarily is a consumer's carton; as such, the requirements are greater for aesthetics than for strength. However, ease of assembly, absence of cracking along edges, and stiffness are important.

The operations associated with the manufacture of folding cartons are printing, die cutting, and gluing. In general, most cartons are printed by flexography, lithography, or rotogravure. Boxboard may be shipped as sheets or as a continuous web, as from a roll. The production of cartons from a continuous web is increasing. In the case of sheet stock, the sheets are fed by hand or by mechanical means to a printing press. One or more colors may be used. After printing, the sheets are fed to a die-cutting press where the carton blanks are cut and creased. The individual carton blanks then are passed through a machine which folds the die-cut blanks into tubes and glues the body closed. The formed tubes are packed flat for shipping. Examples of folding cartons are the toothpaste container, cereal box, butter carton, doughnut box, tack box, and milk carton.

Fiber Cans and Tubes. The basic material used for fiber tubes and cans is a bending board. The body of a fiber can usually is of paperboard and the ends usually are of metal, paperboard, or plastic. The construction of the body may be one of three general types: spiral-wound tubes and cans, convolutely wound tubes and cans, or laminated or lap-seam cans.

Corrugated and Solid Fiber Boxes. Corrugated and solid fiber boxes are used primarily as shipping containers. Both types of containers are made from several layers of paperboard, normally referred to as combined board. Container board is the material from which the combined board is fabricated. Although both types of boxes serve the same general purpose, eg, in the handling, storage, and transportation of commodities, they differ markedly in their manufacture, structure, and performance.

Corrugated board is characterized by its cellular structure which imparts high compressive strength at a relatively low weight. This board usually consists of a corrugated layer with a liner glued to both sides. A relatively thin web of paperboard, ie, the corrugating medium, is passed between two fluted metal rolls to form the corrugation. The facings or liners of high strength container board or linerboard are glued to the tops of the flutes, thereby encasing the corrugated medium on both sides. The various steps take place sequentially: fluting of the central ply or corrugating medium; adhering the liner to one side of the fluted medium by means of aqueous adhesives; adding the other liner or facing by means of adhesives; curing the adhesive and bonding the second liner by passing the formed board over flat, steam-heated plates; and scoring to define flaps and cutting the box blanks to the desired size.

Corrugated board, which is manufactured in this way, is referred to as double-faced corrugated board. Double-wall corrugated board is made by combining two fluted corrugating mediums with a central liner and then adding two outer liner facings. Triple-wall corrugated board is made by combining three corrugated mediums with two inner liners and two outer facings.

The blanks delivered at the end of the corrugator are passed to a printer–slotter, where they are printed using soft rubber dies and where the body scores and slots are introduced into the board. The body scores are similar to the flap scores except that they are parallel to the flutes of the corrugating medium, and thus form the vertical edges of the box when it is assembled. The slotting permits the side flaps to fold over the end flaps. The flaps form the top and bottom of the box. After the printing and slotting operations, the blanks are formed into a flat tube by taping with paper or cloth tape, stitching with metal staples, or gluing the two ends of the box. The resultant flat tubes are bundled or palletized for shipment. The final box is set up by folding and gluing, taping, or stapling the top and bottom flaps.

Variations in construction are made by using different weight facings and different flutes. Conventional fluted rolls are designated A, B, C, and E flutes. The F flute is a more recent addition aimed at the creation of packaging for software and fast foods where more expensive solid bleached board would otherwise be used. They differ in height and in the number of flutes per length of board. The dimensions listed below are approximate, as they vary slightly from manufacturer to manufacturer:

Flute type	Height, not including facings, cm	Flutes m
A flute	0.48	118 ± 10
B flute	0.24	164 ± 10
C flute	0.36	138 ± 10
E flute	0.12	308 ± 13
F flute	0.08	415 ± 15

Cold corrugating is being developed and is expected to be a popular process (33).

Solid fiber combined board consists of numerous bonded plies of container board which form a solid board of high strength. It is much heavier in weight for a given thickness than corrugated board. Solid fiber combined board is made by passing two or more webs or plies of paperboard between a number of sets of press rolls. Adhesive is applied to each ply before it passes through the press nips. In general, solid fiber combined board is made of two to five plies, with three- and four-ply board being most common. The combined weight of the component plies, exclusive of adhesive, is 556–1758 g m² (114–360 lb 1000 ft²). In the combining operation it is customary to join the central plies first and then work outward so that the outer plies are applied last. It also is common to use a poor grade of paperboard, eg, chipboard, in the central plies and a strong linerboard as the outside facings or liners. The subsequent operations are similar to those described for corrugated boxes. As in the case of corrugated boxes, different constructions can be obtained by varying the components, number of plies, and caliper of the solid fiberboard. The adhesives used in the manufacture of corrugated and solid fiber combined board usually are starch or silicate, except where water resistance is required, in which case starch–resin, modified silicate, and resin emulsions may be used.

Flexible Containers.

Paper Bags. There are many types of paper bags, which differ in shape, style, and number of plies, eg, single-, double-, and multiwall. There are a number of bagmaking machines which cut, fold, and glue bags from a continuous web of paper. A variety of papers are used, including bleached and unbleached. The type used depends on the requirements of the product to be packaged. Perhaps the most common are the brown kraft bags which include the grocery and multiwall bags. Kraft bags are used for strength and frequently are of double thickness.

There are four types of single-wall bags. (1) Flat bags are the simplest in construction and the least expensive to make. They have a single lengthwise seam and the bottom is folded under and glued. (2) Satchel-bottom bags provide a flat base when filled. (3) Square-bottom bags, eg, grocery bags, have bottoms similar to flat bags, but have bellows folds at the sides to reduce the width of the closed bag without reducing capacity. (4) Automatic self-opening bags combine the desirable features of the other types of bags. When filled, they form a neat, squared-up package with a stable base and a center or side seam.

Multiwall Shipping Sacks. The construction of multiwall-paper shipping sacks is dictated by the nature of the contents and the shipping and storage conditions. They are used primarily for the packaging of materials that need no protection against compressive forces. Their principal function is to contain the contents and to protect them from contamination. One of the primary requirements of multiwall sack paper is the ability to absorb energy without rupturing (see PACKAGING, CONTAINERS FOR INDUSTRIAL MATERIALS).

Shipping sacks are made from one to six plies of high quality paper and often in combination with special coatings, laminations, or films. In a multiwall sack, each ply or wall is fabricated as a tube and is arranged one within the other, so that each layer bears its share of any applied or induced stress. Better performance is obtained against shock or impact by the use of several plies of relatively lightweight papers than by the use of fewer plies of heavier weight papers. However, the latter generally is considered to be more effective against externally applied point stresses, eg, protruding nails or broken floorboards. The average heavy-duty, multiwall sack is constructed of a number of plies of paper with basis weights of 65–114 g m² (40–70 lb 3000 ft²). The most frequently used basis weights are 65, 81, and 98 g m² (40, 50, and 60 lb 3000 ft²). Papers of heavier weight generally are used in single- and double-ply, pasted-type shipping sacks.

The shipment of many commodities may require special barriers on the sacks to impart resistance against liquids or vapors. Other treatments are used to provide grease resistance, acid resistance, and scuff resistance. Special coatings are used in sacks for packing commodities, eg, synthetic rubbers, asphalts, waxes (qv), and resins, to prevent the contents from sticking to the paper.

Whereas the production of paper sacks has been in decline due to competition from plastics, there are sectors where the ability to manufacture bags from recycled materials and to provide biodegradability favor paper. Bags for fast foods and agricultural products are examples where paper competes favorably against plastic substitutes. Similarly, styrofoam cups and clamshells were common in fast food franchises, however these products have largely been supplanted by paper and paperboard products.

BIBLIOGRAPHY

"Paper" in *ECT* 1st ed., Vol. 9, pp. 812–842, by H. F. Lewis and co-workers, The Institute of Paper Chemistry; "Paper Coatings, Inorganic" in *ECT* 1st ed., Vol. 9, pp. 842–858, by G. Haywood, West Virginia Pulp and Paper Co.; "Paper Coatings, Organic" in *ECT* 1st ed., Vol. 9, pp. 858–867, by P. H. Yoder, Pyroxylin Products, Inc.; "Paper" in *ECT* 2nd ed., Vol. 14, pp. 494–532, by R. P. Whitney and co-workers, The Institute of Paper Chemistry; in *ECT* 3rd ed., Vol. 16, pp. 768–803, by G. A. Baum and co-workers, The Institute of Paper Chemistry.

1. J. N. McGovern, *Pulp Pap.* **52**(9), 112 (1978).
2. D. Hunter, *Papermaking*, 2nd ed., Alfred A. Knopf, Inc., New York, 1947.
3. D. C. Smith, *History of Papermaking in the United States (1691–1969)*, Lockwood Publishing Co., New York, 1970.
4. *FAO World Pulp and Paper Consumption Outlook—Phase H*, Industry Working Party for the Forestry Department of the Food and Agricultural Organization of the United Nations, New York, Sept. 1977.
5. R. W. Mann, G. A. Baum, and C. C. Habeger, *Tappi* **63**(2), 163 (1980).
6. R. H. Mark, *Handbook of Physical and Mechanical Testing of Paper and Paperboard*, Marcel Dekker, Inc., New York.
7. K. W. Britt, *Handbook of Pulp and Paper Technology*, 2nd ed., Van Nostrand Reinhold Publishing Corp., New York, 1970; J. P. Casey, *Pulp and Paper*

Chemistry and Chemical Technology, 3rd ed., Vol. I, Wiley-Interscience, New York, 1980.

8. J. d'A. Clark, *Pulp Technology and Treatment for Paper*, Miller Freeman Publications, Inc., San Francisco, Calif., 1978.

9. R. G. MacDonald and J. N. Franklin, *Pulp and Paper Manufacture*, 2nd ed., Vol. 3, McGraw-Hill Book Co., New York, 1970.

10. **U.S. Pat. 1,873,199** (Aug. 23, 1932), J. D. Haskell; U.S. Pat. 1,985,569 (Dec. 25, 1934), J. D. Haskell; U.S. Pat. 1,960,753 (May 29, 1934), G. P. Prathee.

11. **U.S. Pat. 864,359** (Aug. 27, 1907), G. D. Claflin, Jr.

12. **U.S. Pat. 2,627,477** (Feb. 3, 1953), W. F. Downey (to Hercules Powder Co.).

13. **U.S. Pat. 3,923,745** (Dec. 2, 1975), D. H. Dumas (to Hercules, Inc.).

14. **U.S. Pat. 3,102,064** (Aug. 27, 1963), O. B. Wurzburg and E. D. Mazzarella (to National Starch and Chemical Corp.).

15. J. F. van Oss and C. J. van Oss, *W'arenkennis en Technol.* **V**, 679.

16. J. H. Pulkowski, *Pulp Paper Can.* **94**(6), T166 (1993).

17. *Pigmented Coating Processes for Paper and Board*, TAPPI Press Book No. 28, Technical Association of the Pulp and Paper Industry, Atlanta, Ga., 1962.

18. *Physical Chemistry of Pigments in Paper Coating*, TAPPI Press Book No. 38, Technical Association of the Pulp and Paper Industry, Atlanta, Ga., 1977.

19. *1994 Statistics on Paper, Paperboard, & Woodpulp*, American Forest & Paper Association, Washington, 1994.

20. B. L. Browning, *Analysis of Paper*, 2nd ed., Marcel Dekker, Inc., New York, 1977; T. E. Conners and S. Banerjee, *Surface Analysis of Paper*, CRC Press, Boca Raton, Fla., 1995.

21. L. R. Schroeder and F. C. Haigh, *Tappi* **62**(10), 103 (1979).

22. S. J. V. Yound, C. Finsterwalder, and J. A. Burke, *J. Assoc. Off. Anal. Chem.* **56**, 957 (1973).

23. I. H. Isenberg, *Pulp and Paper Microscopy*, 3rd ed., The Institute of Paper Chemistry, Appleton, Wis., 1967.

24. J. Quarles, *Federal Regulations of New Industrial Plants*, Morgan, Lewis and Brochius, Washington, D.C., 1979.

25. J. G. Strange, *The Paper Industry—A Clinical Study*, Graphic Communications Center, Inc., Appleton, Wis., 1977.

26. H. S. Dugal, "Environmental Laws and Their Impact on Mill Processes," paper presented at *The Third Symposium on Corrosion in Pulp and Paper Industry*, May 1980.

27. *Environmental Pollution Control, Pulp and Paper Industry, Part 1—Air*, EPA 625 7-76-001, U.S. EPA, Washington, D.C., Oct. 1976.

28. *Process and Design Manual for Sludge Treatment and Disposal*, EPA 625 1-74-006, U.S. EPA, Washington, D.C., Oct. 1974; D. Marshall, *South. Pulp Pap. Manul*, 19 (Dec. 1977).

29. J. J. Reinhardt and D. F. Kolberg, *Pulp Pap.*, 128 (Oct. 1978).

30. *Pap. Trade J.*, 34 (May 15, 1979).

31. *Process Design Manual for Sludge Treatment and Disposal*, EPA 625 1 74-006, U.S. EPA, Washington, D.C., Oct. 1974.

32. R. Patrick and D. M. H. Martin, *Biological Surveys and Biological Monitoring in Fresh Waters*, Academy of Natural Sciences, 1974.

33. J. M. Itellawell, *Biological Surveillance of Rivers*, Water Research Centre, Herts, U.K., 1978.

General References

J. A. Bristow and P. Kolseth, *Paper Structure and Properties*, Marcel Dekker, Inc., New York, 1986.

D. Gunderson, *Paper Timber* **74**(5), 412 (1992).

H. F. Rance, *Handbook of Paper Science*, Elsevier, Amsterdam, the Netherlands, 1980.

Tappi Press Books, Technical Association of the Pulp and Paper Industry, Atlanta, Ga.

ASTM Standards, American Society of Testing Materials, Philadelphia, Pa.

M. Bruce Lyne
Paper Science, International Paper

PAPERMAKING ADDITIVES

In papermaking, chemicals can be added either to the pulp slurry prior to sheet formation, ie, internal or wet-end addition, or to the resulting sheet after complete or partial drying, ie, surface or dry-end addition. The method chosen depends on retention and the desired effect. For example, strength additives usually are added internally if uniform strength throughout the sheet is wanted, but they are applied to the surface if the need is for increased surface strength. If an additive cannot be retained efficiently from a dilute pulp slurry, then it is better to apply it to the surface of the sheet.

Papermaking additives can be categorized either as process additives or as functional additives. Process additives are materials that improve the operation of the paper machine, such as retention and drainage aids, biocides, dispersants, and defoamers; they are primarily added at the wet end of the paper machine. Functional additives are materials that enhance or alter specific properties of the paper product, such as fillers (qv), sizing agents, dyes, optical brighteners, and wet- and dry-strength additives; they may be added internally or to the surface of the sheet.

Environmental constraints on the paper industry have resulted in drastic processing changes, primarily because very large amounts of water are used to produce paper. If chemical additives are not efficiently retained in the sheet, in addition to losing the value of these materials, the concentration of these materials increases in the white-water system (the water drained and pressed from the fibers as the sheet is being formed). As paper mills become more closed, ie, as they reuse the white water many times over before treating and discharging it, retention of papermaking additives becomes more critical because the unretained materials also negatively impact the performance of newly added materials as well as the finished sheet. Retention of papermaking additives also becomes more difficult with the increased use of recycled waste fiber and the closing of the water system, because the conductivity and amount of soluble materials increase under these conditions. Some additives become electrostatically neutralized by these materials; others are forced to compete for binding sites on the cellulose fibers.

The U.S. Environmental Protection Agency (EPA) proposed the "cluster rule" for the paper industry in December 1993; it provided detailed and comprehensive guidelines regarding discharges of harmful materials to the air, water, or soil (sludge) (1). The revised Clean Air Act (1990) also identified hazardous air pollutants whose discharge is stringently regulated (2). Numerous regional, state, local, and foreign national regulations exist concerning emissions to air, and discharges to water and sludge. OSHA workplace regulations may have also altered the additive process and the choice of additives.

In addition, many grades of paper and paperboard are used in direct or indirect contact with foods. Thus, many mills only use paper chemicals that have been cleared for use by the U.S. Food and Drug Administration (FDA) (3), so that it is not necessary to segregate machine broke (off-grade paper and edge clippings that are reclaimed for their fiber value) and white water. Most of the chemicals discussed in this article are approved by the FDA for use in paper and paperboard that are intended for applications in food processing and packaging. However, there are various restrictions on both the specific functional uses and amounts of paper chemical additives which can be used, so the FDA status should be confirmed by the supplier before use.

It is also important to study the interactions of papermaking additives (4) in the paper machine water system; some additives act synergistically, so that the performance of each is enhanced by the presence of the other. However, some additives have a negative impact on the performance of other additives, or on other desirable paper properties. Thus, optimization of the addition points and usage rates of the entire additive system is necessary in order to maximize performance of the chemical additives and the paper sheet properties, and to minimize cost and negative interactions both on the paper machine and in the white-water system. This is especially true as unanticipated additives enter the wet end of the paper machine from recycled furnishes, including coated broke (5).

Lists of the manufacturers of each type of product used by the paper industry are available (6). Surveys of chemical suppliers for the U.S. and Canadian pulp and paper industries have been published (7).

Process Aids

Process aids, which improve the operation of the paper machine, include retention aids, ie, coagulants and flocculants; drainage aids; effluent treatment for fiber and filler recovery; formation aids; defoamers; wet-web strength additives, pitch-control agents; creping aids; and biocides and slimicides. The first three are phenomenologically related: they act by agglomerating filler particles, fines, or fibers with themselves or with each other. Agglomeration occurs as a result of electrostatic attraction (ie, cationic polymers or alum, known as charge modifiers, neutralized by anionic fibers, fines, or inorganic fillers) or hydrophobic interaction (ie, sparingly water-soluble pitch or defoamer components adhering to dispersed filler particles). Thus, in one papermaking system, an additive may increase first-pass retention, whereas at another mill it would act more as a drainage or dewatering aid (8).

Retention and drainage additives are vital to the use of recycled fibers. Papermakers consider recycled fibers to behave like virgin fines, while recycled fines behave like filler. Drainage on the paper machine can be impeded and first-pass retention reduced by the use of recycled fiber (9). Additionally, the negative impact of contaminants found in recycled fibers can be minimized by the appropriate use of dispersants and other pitch-control additives.

Retention Aids. Opacifying filler particles usually are ca 0.3–5 µm in diameter and are much smaller than most pulp fibers. Therefore, they are not effectively retained by filtration through the pulp mat as it forms on the machine. In an aqueous suspension, most fillers, like most paper-pulp fibers and fines, develop a negative surface charge which prevents coflocculation of fillers with fines. Retention aids encourage coflocculation by two mechanisms: they neutralize the negative charges on fillers, fibers, or fines so that van der Waals forces can hold them together, and they form molecular bridges between two particles to which they are adsorbed.

Molecular weights of the charge-biasing polymeric retention aids are ca 10³–10⁵. These aids usually contain amine or quaternary ammonium groups and include condensation polymers, eg, diethylenetriamine–adipic acid polyamide [25085-20-5], which are treated with epichlorohydrin [106-89-8] and ammonia—or dimethylamine—epichlorohydrin condensates; ring-opening polymers, eg, polyethylenimine [9002-98-6]; and addition polymers, eg, poly(dimethyldiallylammonium chloride) [26062-79-3] or copolymers of aminoethyl acrylates. These products are supplied as 10–35% aqueous solutions.

The simplest charge-biasing agent is an alum [10043-01-3] which is hydrated aluminum sulfate. The alum referred to in this article is papermakers alum, not aluminum potassium sulfate (see ALUMINUM COMPOUNDS, ALUMINUM SULFATE AND ALUMS). Alum is used to control sheet formation through fiber flocculation, to improve drainage, and to precipitate rosin size. It is used less as a sole retention aid than as an adjunct to other retention aids, especially the bridging types. Alum facilitates adsorption of bridging polymers, eg, by neutralizing negative charges on water-soluble and particulate impurities.

Molecular weights of polymers that function as bridging agents between particles are ca 10⁴–10⁷. Ionic copolymers of acrylamide are the most significant commercially (see ACRYLAMIDE POLYMERS). Cationic comonomers include (2-methacryloyloxyethyl)trimethylammonium salts, diethylaminoethyl methacrylate [105-16-8], and dimethyldiallylammonium chloride [7398-69-8]; anionic comonomers include acrylic acid [79-10-7] and its salts. Both types of polyacrylamides, but especially the anionic, can be more effective in the presence of alum (10,11). Polyethylenimine and vinylpyridine polymers, eg, poly(1,2-dimethyl-5-vinylpyridinium methylsulfate) [27056-62-8] are effective but are used less frequently.

The high molecular weight bridging polymers are more effective retention aids than the lower molecular weight, charge-biasing materials. However, the former cannot be shipped economically, except as solids, because of their high solution viscosities. These solid products must be dissolved with care to prevent the formation of slowly dissolving lumps. Invert emulsions, ie, aqueous polymer solutions that are emulsified in a solvent, and logs of water-swollen, super-high molecular weight (>10⁷) polymer have been introduced to circumvent problems of solution makeup. The lower molecular weight, charge-biasing polymers are sold as solutions and can be diluted easily.

Synergistic improvements in filler retention have been achieved through the use of combinations of additives, in which addition of a low molecular weight cationic polymer, often referred to as a coagulant, is followed by that of a high molecular weight anionic polymer, or flocculant (11). These probably function through a combination of charge biasing and bridging. Although they provide very high retention of fillers and fines, these combinations can disrupt formation by overflocculating the fibers (12). Such overflocculation can be compensated for by redispersing the flocs with vigorous agitation of the stock just before it passes onto the forming wire (13). It has been found that the addition of a high molecular weight cationic flocculant prior to a point of shear, followed by a medium-to-high molecular weight anionic flocculant, improves fines and filler retention without impeding drainage (14,15).

Another type of dual-retention system has emerged; in these systems, a synergistic combination of a small, inorganic particle and a polymer enhances retention. These microparticulate retention systems are composed of a high molecular weight cationic flocculant and an anionic particle. The most common of these systems consist of combinations of cationic starch with anionic aluminum hydroxide or colloidal silica, or high molecular weight cationic polyacrylamides with bentonite or colloidal silica (16). The cationic polymer flocculates the fines, fillers, and fibers. The anionic microparticle, added after shearing the suspension, is believed to reflocculate or strengthen the previous flocs, owing to its high specific surface and anionic charge. The anionic microparticle, because of its small size, appears to increase the density of the flocs, thus reducing their size (17). Unlike purely polymer-based dual retention systems, the microparticle systems are reversible, in that the floc system reverts to its original structure once shear is removed (18).

The percent retention of mineral fillers usually is measured directly by ash analysis of the sheet, with appropriate corrections for weight loss of the particular filler during ashing. During production, it can be calculated by difference from the solids content of the white water. The opacifying power of a filler depends on its distribution in the sheet. A finely divided filler that is distributed uniformly on fibers opacifies paper more effectively than the same amount of filler that is distributed in fewer, larger aggregates. The scattering coefficient, a measure of the optical efficiency of the retained filler, is calculated from the basis weight, ie, the sheet weight per unit area, and reflectance of sheets against white and dark backgrounds using the Kubelka-Munk equation (19) or derived forms (20–22). The efficiency of a retention aid depends on the amount added, the shear, and the molecular weight. An overdose of retention aid can redisperse the filler.

Retention can be studied as a function of both dose and shear in the dynamic drainage jar (DDJ), in which a variable-speed agitator exposes the stock to controlled shear and prevents mat formation during drainage (23). Thus, the shear dependence of retention, which results from electrokinetic and colloid effects, can be studied, as distinct from retention resulting from filtration on a fiber mat. The degree of shear on a given paper machine can be quantified as an rpm-equivalent in the drainage jar by varying the agitator speed in the DDJ test for a given stock furnish until the percent retention equals that on the paper machine.

Salts, eg, alum or calcium chloride [10043-52-4], and cationic polyacrylamides are effective retention aids in bleached and unbleached kraft pulp.

Formation Aids. The repulsion of negatively charged fibers in water is not always sufficient to prevent all flocculation, which can result in uneven fiber density in paper. Dispersants can prevent flocculation prior to immobilization of the wet fiber mat on the wire. Polyacrylamide [9003-05-8], poly(ethylene oxide) [25322-68-3], and natural gums, eg, guar gum [9000-30-0] and locust bean gum [9000-40-2], promote even fiber distribution. High molecular weight anionic polymers, such as polyacrylates, lignin sulfonates, and naphthalene sulfonates, can also act as dispersants and enhance sheet formation, but they may impede drainage in some systems (24).

Drainage Aids. The break-even point in a typical paper mill is high, ie, ca 85% of capacity. Therefore, small increases in productivity can cause large increases in profits and can postpone the massive capital investment, ie, >\$500 × 10⁶ required to build a new paper machine. For machines where dewatering on the Fourdrinier wire is the slow step, dewatering or drainage aids can effect faster machine operation and, therefore, can increase production (25). However, a drainage aid can improve drainage on one section of the machine and impede it on another (8). For example, high molecular weight cationic flocculants can accelerate initial percolation of water on the wire by flocculating the pulp and allowing channelling, but further dewatering on the machine suction boxes is impeded by air leakage through the channels, which reduces the pressure gradient (26). Thus, an ideal drainage aid alters the surface properties of the fibers so that they hold less water, and limits the amount of fiber flocculation. This balance of properties can be achieved by combining a drainage aid with sufficient agitation to redisperse the fiber flocs immediately prior to sheet formation.

Defoamers. Foam is a common problem in papermaking systems (27). It is caused by surface-active agents which are present in the pulp slurry or in the chemical additives. In addition, partially hydrophobic solid materials can function as foam stabilizers. Foam can exist as surface foam or as a combination of surface foam and entrained air bubbles. Surface foam usually can be removed by water or steam showers and causes few problems. Entrained air bubbles, however, can slow drainage of the stock and hence reduce machine speed. Another serious effect is the formation of translucent circular spots in the finished sheet caused by permanently entrained air.

Defoamers (qv) are available in several forms, composed of many different materials. Historically, paste and solid defoamers were used extensively. Composed of fatty acids, fatty amides, fatty alcohols, emulsifiers (and mineral oil [8012-95-1] in the high solids paste emulsions), these defoamers required emulsification (brick) or dilution (paste) before use. Liquid defoamers have become the preferred form, insofar as concern about handling and overuse have been overcome.

Some liquid defoamers are preemulsified relatives of paste defoamers. In addition to the fatty components mentioned above, kerosene [8008-20-6] or an organic cosolvent such as 2-propanol have been used to enhance stability of the oil—water emulsion and the solubility of the defoamer's active ingredients. These cosolvents are used less frequently as concerns increase about volatile organic emissions (VOCs) from the paper machine. Additionally, the use of ultrapure mineral oil in defoamers has become commonplace. Concern about the creation of 2,3,7,8-tetrachlorodibenzodioxin (TCDD) and 2,3,7,8-tetrachlorodibenzofuran (TCDF) in the pulping process has led to the discovery of unchlorinated precursor molecules, especially in recycled mineral oil and other organic cosolvents used in defoamer formulations (28). In 1995 the mineral oil that is used is essentially free of dibenzodioxin and dibenzofuran. In addition, owing to both the concern about these oils and the fluctuating cost of raw materials, the trend in paper machine defoamers is toward water-based defoamers (29).

The defoamer formulations mentioned so far consist of fairly inexpensive raw materials, but several more costly defoaming materials have come into use in paper mills. Hydrophobicized silica particles are useful in some emulsion formulations. Silicone solutions and emulsions are very effective in eliminating foam in paper machine water systems. The silica- or silicone-based defoamers have higher activity, which somewhat compensates for the higher cost, but care must be taken to prevent overuse.

Wet-Web Strength Additives. When the wet sheet is transferred from the forming wire to the press section of the paper machine, it still contains 60–75% water. If there is inadequate sheet strength, breaks occur, resulting in machine downtime and lost production. A number of water-soluble chemical additives have been shown to enhance wet-web strength, including synthetic, eg, anionic polyacrylamides, and natural, eg, locust bean and guar gum, polymers. Several modified natural polymers have also shown promise, including chitosan (30), blocked reactive group starches with acetal protected aldehydes (31), and cationic aldehyde starches produced from them (32). These have replaced the previously successful dialdehyde starch which is made in Japan (31). Since these additives improve wet-web strength indirectly by improving formation and accelerating water removal from the forming sheet, other formation and drainage aids may also improve wet-web strength.

Pitch Control Agents. Sticky pitch deposits may form at various locations at the wet end of the paper machine. If these deposits enter the system they may be incorporated in the sheet, which could result in sheet breakage or off-color spots. Historically, the main source of pitch in a papermaking system was the natural wood resins that remain in the virgin pulp. In addition, with the use of increased levels of recycled fibers as a significant portion of the papermaking furnish, other sticky materials, eg, resins, adhesives, waxes, inorganic fillers, and latexes called "white pitch" and "stickies," which are especially troublesome when utilizing coated broke (33), may contribute to pitch problems. Pitch problems can be minimized by more thorough pulp washing and screening, which, however, may not be practical, insofar as the water systems in paper mills are increasingly more closed.

The most widely used pitch control method is the addition of pitch dispersants, which can be either organic, ie, typically anionic polymers such as naphthalene sulfonates, ligninsulfonates, and polyacrylates (33,34), or inorganic, ie, typically clay or talc. The polymers maintain the pitch as a fine dispersion in the pulp, preventing agglomeration and potential deposition on the paper machine or the sheet. When talc, clay, or other adsorbent fillers are added to the furnish, moderate amounts of pitch can adsorb on these materials, producing a nontacky solid that can be retained in the sheet.

Low molecular cationic polymers or alum can also be used to flocculate pitch, ie, bind up the pitch so that it is retained in the sheet, to minimize pitch deposition on machine surfaces and fabrics (35,36). Alum is used commonly in newsprint operations (34). The addition of a nonionic surfactant with a hydrocarbon solvent to the wet end has shown some utility in preventing deposits of adhesive recycled furnish contaminants from forming on the paper

machine (37).

Effluent Treatment. Like the mineral- and pigment-processing industries, the paper industry must minimize the amount of suspended solids in mill effluents. Save-alls, which are specific to the paper industry, are designed to recover paper fines and filler particles for recycle to the paper machine. Flocculants are used both in the paper machine furnish to increase retention of fines and fillers in the sheet (38) and in save-alls to help recover finely divided solids, both organic and inorganic, from the white water. Alum is used widely as a coagulant for effluent treatment in paper mills. Most of the low and high molecular weight polymers, which serve as retention aids, also can be used as coagulants and flocculants (see FLOCCULATING AGENTS).

Slimicides and Biocides. Most paper machine systems are ideal environments for the growth of slime-forming bacteria and fungi because of the many nutrients which are present, eg, hemicelluloses, starches, and other organic additives, and because temperatures often are optimum for microorganism growth. The most effective way to prevent slime formation is to maintain cleanliness in the entire papermaking system. However, in most mills, cleanliness is not sufficient to completely prevent slime formation. Therefore, slimicides and biocides are frequently added at various points in the papermaking process. This is of crucial importance when the paper is produced for food contact or medical applications.

There are two categories of biocides in use in paper mill systems: oxidizing biocides, including chlorine, hypochlorite, hypobromous acid, and chlorine dioxide, and nonoxidizing biocides, ie, methylene bithiocyanate, carbamates, and quaternary ammonium compounds (39). When alkaline paper is produced, biological growth can be a serious problem; maintaining a residual free chlorine level (0.3–0.5 ppm) in the process water has been recommended (33). Because of the specificity of the effectiveness of these materials (40), it is best for paper mill personnel to work closely with biocide suppliers to select the best control agent for a given situation (see INDUSTRIAL ANTIMICROBIAL AGENTS).

Creping Aids. Disposable sanitary papers, which are used in napkins, towels, and facial and toilet tissue, should absorb water. Although surface-active agents and some strength resins can improve absorbency, the more prevalent method involves mechanically creping the sheet. Mechanical creping simultaneously improves softness and absorbency. Creping is achieved by peeling the sheet from a steel drier roll with a sharpened doctor blade, which is maintained at an angle to the surface of the roll. Successful creping depends on the angle and design of the doctor blade and the degree of adhesion of the sheet to the roll. If the sheet adheres excessively, too much dry strength is lost as the sheet is chiseled from the roll. Principal creping adhesives include animal glues, starch, neutral-cure wet-strength resins, specialized polyamines, and high molecular weight retention aids. Among the agents that can be added to facilitate release from the roll are emulsified paraffin oil [8012-95-1], silicone oils, or poly(ethylene glycol)s (41). Proprietary combinations of polyamide—epichlorohydrin resins and release agents sometimes are used to tailor the proper balance of adhesion to release for a given pulp on a given machine, at a given sheet moisture content. Residual hemicellulose in the pulp affects adhesion significantly.

Functional Internal Additives

Sizing Agents. Paper sizing provides paper and paperboard with resistance to wetting by liquids. Because aqueous fluids, eg, ink, water, and milk, generally are the liquids of concern, the usual purpose of sizing is to produce water repellency. Most grades of paper that are to be printed or written on are sized so that the ink does not spread laterally or feather. For other uses, it is necessary to size paper to prevent penetration by liquids, eg, in paper cups, milk cartons, and packaging papers and boards (see also WATERPROOFING AND WATER OIL REPELLENCY).

The contact angle formed between water and the paper surface is the primary factor determining the extent of wetting, as shown in Figure 1 (42). Well-sized paper has an initial contact angle of 91–100°, which permits only limited wetting and spreading; there is no tendency for water to penetrate through pores. However, if the contact angle is considerably less than 90°, water wets, spreads, and penetrates the sheet quite rapidly. When properly oriented on the fiber surfaces, purely hydrophobic materials such as wax, and amphipathic, ie, polar—nonpolar, materials such as rosin[8050-09-7] provide a low surface energy coating which gives the high contact angle necessary for sizing.

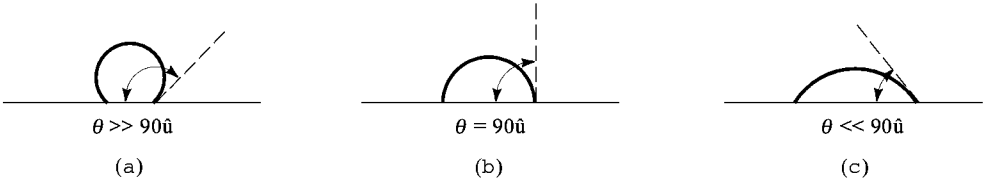


Fig. 1. Physical phenomena governed by contact angle: (a) extremely limited wetting and spreading, tendency to retract, does not penetrate; (b) limited wetting and spreading, no tendency to penetrate; (c) extensive wetting and spreading, strong tendency to penetrate.

The rate of flow of a liquid through a very thin tube or capillary is represented in Figure 2 by the Washburn equation, which combines the equation representing the natural driving force for fluid movement in a capillary tube with the Poiseuille equation for laminar flow through a tube (43). Two of the five parameters that govern the rate of flow, ie, surface tension and viscosity of the penetrating liquid, are determined by the customer's needs, ie, the liquid whose movement the sized paper product should resist. Two more, ie, the radius and length of the pores, are governed by the papermaker's products, ie, the basis weight, bulk density, and porosity of the sheet. Thus, the chemical supplier can vary only the composition of the sizing agent to give the contact angle needed to produce the sizing that is required by the paper producer and consumer.

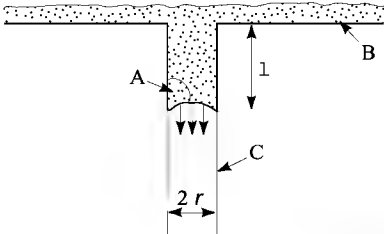


Fig. 2. Liquid flow-through capillary (Washburn equation). Time rate of penetration = $dl/dt = 1/4 \sqrt{\gamma/\eta} \times [r/l] \times \cos\theta$, where γ = surface tension and η = viscosity. A, contact angle θ between liquid and capillary wall; B, penetrating liquid; C, partially filled capillary, r = radius, and l = length already filled.

Rosin-Based Sizes. These sizes are normally provided in one of three forms: paste, liquid and emulsion. Paste rosin sizes are supplied as viscous pastes containing 60–80% solids. These sizes may contain unmodified or fortified rosin that has reacted (ie, been fortified) with either maleic anhydride [108-31-6] or fumaric acid [110-17-8] (see Fig. 3). In either case, the unmodified or fortified rosin is treated with aqueous alkali so that the degree of neutralization, ie, saponification, varies from 75–100% depending on the physical state desired for the commercial product. Before use, the paste size must be converted to a stable, dilute rosin size emulsion by careful sequential dilution with warm water followed by cold water, with good agitation.

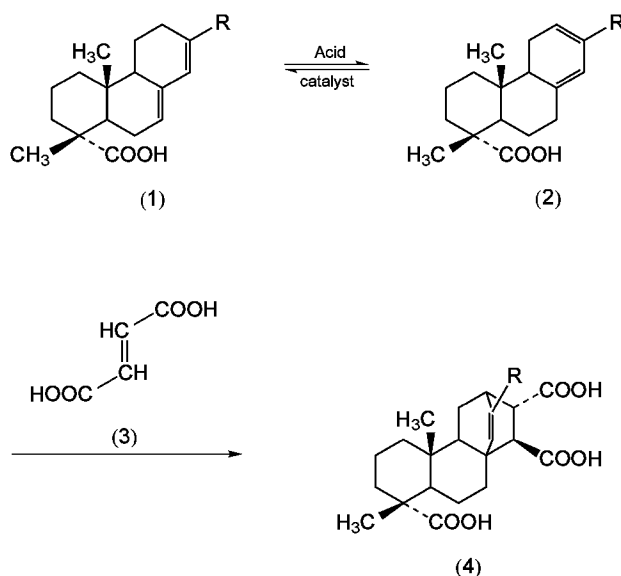


Fig. 3. Resin acids in rosin sizes, $R = \text{CH}(\text{CH}_3)_2$. The resin acids are represented here as abietic acid [514-10-3] (1) and levopimaric acid [79-54-9] (2). In rosin there are other isomers and disproportionation products. The product of reaction with fumaric acid (3) is levopimaric acid—fumaric acid adduct [77942-79-1] (4).

Liquid rosin sizes are based on more highly fortified rosin than those used in paste sizes. The highly fortified rosin is then neutralized 100% with aqueous alkali. These sizes are very low viscosity liquids typically in the 45–55° solids range. The advantages that liquid sizes have over paste sizes lie in their handling and sizing efficiency. Due to its lower viscosity, liquid size can be used directly without emulsification or can be simply diluted with cold water, thus eliminating the need for heated storage and emulsification equipment. Since they are highly fortified, liquid sizes generally require lower usage rates than paste sizes to achieve a specific degree of sizing.

Fortified rosin can also be converted to high free-rosin emulsions by using various stabilizers. Typically, these are 35° solids emulsions, which exhibit excellent stability in relation to storage and mechanical action, such as is found in transfer pumps. Generally, fortified rosin emulsions are more efficient sizes than their soap-based liquid or paste counterparts.

Rosin sizing usually involves the addition of dilute aqueous solutions or dispersions of rosin soap size and alum to a pulp slurry (44–46). Although beater addition of either coreactant is permissible, addition of both before final pulp refining is unwise because subsequently exposed cellulose surfaces may not be properly sized. The size and alum should be added sufficiently early to provide uniform distribution in the slurry, and adequate time for the formation and retention of aluminum resins, commonly referred to as size precipitate. Free rosin emulsion sizes, however, do not react to a significant degree with alum in the pulp slurry, and addition of a cationic starch or resin is recommended to maximize retention of size to fiber. Subsequent reaction with aluminum occurs principally in the machine drier sections (47).

These rosin-based sizes, whether paste, liquid, or emulsions, can be used to size all grades of paper that are produced at acid pH. The latter include bleached or unbleached kraft linerboard and bag paper, bleached printing and writing grades, and cylinder board. In addition, polyaluminum compounds have been used in place of alum, most notably, polyaluminum chloride (48), which can reduce barium deposits where these have been a problem. The barium chloride by-product is more water-soluble than barium sulfate. Other polyaluminum compounds such as polyhydroxylated forms of alum and polyaluminum silicosulfate have been evaluated as alum replacements.

Rosin sizes, most notably emulsion sizes, have gained increased usage in grades of paper produced at neutral pH. The principal advantage of neutral papermaking is that it allows increased usage of recycled white paper and machine-coated broke that contains calcium carbonate filler, and therefore would require operating at a higher pH to avoid the sizing and machine runnability problems associated with soluble calcium and volatile carbonate, which produces extensive foaming at an acid pH (49).

Cellulose-Reactive Sizes. There are two principal types of synthetic paper sizes that owe their effectiveness to the hydrophobic size molecule actually forming a covalent bond with the cellulose (50). The first cellulose-reactive sizes were based on hydrolytically stable emulsions of alkylketene dimers (AKD) of long-chain fatty acids. More recently, cationic AKD emulsions have been developed that produce more rapid on-machine sizing during drying than the original unmodified products (51). The other significant cellulose-reactive internal sizing additive is alkenyl succinic anhydride (ASA), made by reaction of a long-chain alkene with maleic anhydride. Since ASA hydrolyzes rapidly in water, it is shipped neat and emulsified on-site (52).

Cellulose-reactive sizes have gained prominence because paper can be sized effectively at neutral to alkaline pH with these additives. There are several advantages to sizing paper under slightly alkaline conditions. Inherently, paper formed under alkaline conditions has greater strength than paper formed under acid conditions. This allows for a greater substitution of lower cost filler for higher cost (but higher strength) fiber. Calcium carbonate is useful as a pigment and filler only on the alkaline side. Because paper that is made under alkaline conditions has greater strength, it can be made from a less highly refined furnish which requires less energy for refining and drying. Another significant advantage of alkaline papermaking is the reduction of inorganic salts in the white water, which reduces the cost of effluent treatment and permits greater closure of the white-water system. Alkaline papermaking can result in increased productivity because of lower maintenance cost than had previously been needed owing to acid wear of machine parts, but the abrasiveness of the calcium carbonate pigments causes increased wear on paper machine synthetic wires and other parts (51). Paper formed at an alkaline pH also has greater permanence than acid paper (ie, its strength is retained better over time), and it has thus become important for archival purposes (53).

Alkaline sizing agents are especially effective in milk-carton board and printing and writing grades that utilize calcium carbonate fillers.

Other Internal Sizing Additives. *Wax emulsions* have been widely used to impart special resistance to functional penetrants, such as oil, grease, and blood. Butcher paper, meat wrap, cable wrap, bleached kraft foodboards, and folding boxboards are all paper grades that have utilized wax emulsions to develop a high degree of sizing at minimum cost. However, the use of AKD sizes and high efficiency, rosin-based emulsion sizes has caused a reduction in the use of wax emulsions. Most wax emulsions that are used for internal addition are stable to mild acid and mild alkali. They are retained in the sheet as a result of the addition of alum, which breaks the emulsion; the wax particles are then retained by filtration. Wax emulsions that can be broken by the addition of acid are used to a lesser extent. Cationic emulsions are used to a limited extent; they are retained by addition of sulfate ions to break the emulsions. In general, wax emulsions are added to the pulp furnish after addition of rosin-based size and alum, and just prior to the paper machine headbox.

Fluorochemicals have been used in the manufacture of oil-resistant paper and paperboard, and other specialty grades. The fluorochemicals that are best suited are characterized by a long perfluorocarbon chain attached to either a functional group or a polymer backbone. Suitable fluorochemicals include fluorochemical chrome complexes, fluorochemical copolymers, and fluorochemical phosphates; the latter are the most commonly used, since they are approved by the FDA for direct food-contact packaging. These products, provided as 33 wt % solids solutions in water with an organic cosolvent, can be retained efficiently by the use of cationic retention aids. Although fluorochemicals provide excellent oil and grease resistance, they do not provide sizing against aqueous penetrants. When both oil and water resistance are desired, the fluorochemical phosphates usually are used with AKD sizes.

Mechanism. The general mechanism of effective sizing involves the following sequential steps. (1) Efficient retention of the sizing agent in the

sheet, preferably as tiny particles. (2) Uniform distribution of sizing agent over fiber surfaces. If the size has not been retained in appropriately small, well-distributed particles, sizing distribution must result from spreading during drying. (3) Firm anchoring to fiber surfaces so that impinging liquids do not overturn the molecules which present a hydrophobic surface to the exterior. Ideally, the sizing agent should be chemically inert to the liquids which are encountered in use and should not adversely affect other paper properties.

Aluminum resinate particles, ie, size precipitate, are attracted to the fiber surfaces because of a difference in charge and thus are retained (45,46,54). In general, the particles of size precipitate are small and are distributed fairly uniformly over the sheet. However, on drying, there is some sintering of the particles which helps to redistribute them on the fibers.

The higher efficiency of fortified rosin sizes is believed to result from the semihydrophilic nature of the rosin adduct molecules, which results in a more dispersed system of particles during size precipitation by alum. Consequently, there is a more uniform distribution of somewhat smaller particles on the sized fibers. This dispersing effect may result from the strong tendency of aluminum to coordinate with organic anions.

During drying, the lower softening point of the size precipitate, which is obtained from high free-rosin emulsion sizes, permits much greater redistribution of the retained sizing agent than is possible with ordinary rosin size and fortified rosin size. It is postulated that the free rosin acids in the retained size precipitate react directly with various aluminum compounds during drying of the sheet to provide the anchoring required for good sizing. After drying, the aluminum resins are immobile below 100°C and are oriented with the hydrophilic carboxyl groups combined with aluminum on the fiber surface, and the hydrophobic bulk of the rosin molecule oriented outwardly.

Synthetic cellulose-reactive sizing agents, such as long-chain alkylketene dimers and alkenyl succinic anhydride, are added to the pulp slurry as emulsions and are retained in the sheet by cationic retention aids (cationic polymers or cationic starch), which also may serve as part or all of the emulsifier system. The retained sizing agent has a low softening point, a property which permits it to distribute uniformly over fiber surfaces during drying. The size also reacts with hydroxyl and carboxyl groups on the cellulose fiber surface to form esters and anhydrides, which firmly anchor it to the paper surface. As the paper dries, there is a competing reaction of the size with water, which renders some of the sizing agent ineffective.

In the production of alkaline paper, a phenomenon called size reversion, ie, paper that possesses the appropriate level of sizing response when produced then loses all or part of that sizing response upon aging, can sometimes be observed. Several factors appear to be at work. The use and overuse of polyamines and other promoters that enhance the on-machine sizing response of AKD can also promote the hydrolysis of AKD as the paper ages, with a potential loss of sizing (55). The use of highly alkaline calcium carbonate filler slurry can cause a similar loss in the sizing response over time, for both AKD- and ASA-containing sheets (55,56). The effect is not seen in paper filled with clay alone, unless the papermaking slurry is made excessively alkaline. With all sizing agents (acid or alkaline), sizing is most efficient if the size is retained on the fiber surfaces, as opposed to the filler surfaces (44,57).

Sizing Tests. A thorough description of sizing tests can be found in Reference 58. No one test is unanimously endorsed or entirely satisfactory. A practical sizing test should either measure end use requirements directly, simulate use conditions, or correlate with use requirements. A reliable test also should be sensitive to small changes in sizing level and be reproducible from one operator to another. In printing and writing grades of paper, end use printability tests are increasingly being utilized, as opposed to continuing to try to correlate size tests which mimic the penetration of ink solutions. One printability test that is increasing in popularity with uncoated free sheet manufacturers is the ink jet printability test. This test can be performed with either a black-and-white or color printer. Key components of the test include show-through, feathering, spreading, resistance to smear when highlighted with both acid and alkaline highlight markers, and color-to-color bleed. This test is a more accurate predictor of paper performance than other size tests.

Dry-Strength Additives. An increasing amount of dry-strength additives is being used. This expanded use is driven largely by the need to utilize higher levels of weaker fiber sources, ie, recycled paper and paperboard. Modified starches are the most widely used, with the cationic and amphoteric starches dominant for wet-end application. Polyacrylamides, both anionic and cationic, are the most important of the synthetic dry-strength additives, and are used in a variety of grades. Vegetable gums, both natural and modified, and sodium carboxymethylcellulose [9004-32-4] (CMC) are also widely used.

Utility. Increased sheet strength through increased fiber refining generally results in an increase in sheet density. This means that increased sheet strength is accompanied by a reduction in density-related properties, such as opacity, porosity and bulk. Dry-strength additives increase sheet strength with little or no change in sheet density (59). This provides for faster drainage during paper formation, and allows for separate optimization of sheet strength and density-related properties. In addition to increased sheet strength as a direct objective, the dry-strength additives permit benefits, that would otherwise be strength-reducing, to be exploited. These include reduced refining energy, higher filler levels, increased use of weaker, less expensive fibers, and reductions in basis weight.

Natural Gums. The mucilages from plant roots and stems, most notably guar gum, locust bean gum, and tamarind gum, may be used by papermakers to improve the strength of paper (60). Galactomannan is the effective natural polymer in the materials. These gums, of which guar is the most commercially available, tend to be irreversibly adsorbed on cellulose pulp fibers and thus augment the hemicelluloses (qv) formed during refining to impart strength to the paper. Cationic and amphoteric guar derivatives have been produced and used as strength additives; in some applications, these derivatives are more effective than the naturally occurring gums. Gums are prepared as aqueous solutions before addition to the pulp slurry for ease of handling. In general, gums are used at 0.1–1.0 wt %, based on the dry pulp fibers (see GUMS).

Starches. Like the natural gums, starches were among the first strength additives used. Both cereal starches, such as corn and wheat, and root starches, eg, potato, are effective, along with numerous types of modified starches. Anionic starches enhanced by carboxylation or phosphorylation, cationic starches produced by addition of tertiary amino or quaternary ammonium cationic groups, amphoteric starches containing both cationic and anionic functionalities, and starches modified with synthetic polymers are all widely used. Cationic starches modified by the addition of an acetal group (blocked aldehyde) can also serve to enhance the dry strength of paper (61).

Like the natural gums, starches need to be cooked in water to form dispersions for addition to the papermaking system. Various techniques have been developed for cooking starches rapidly (see STARCH). In general, anionic starches are used with alum, which aids in starch retention. The cationic and usually the amphoteric starches are self-retaining.

Starches used to increase the internal strength of paper as measured by tensile, edge crush resistance, Mullen burst, and Scott bond strength are added at the wet end, whereas starch used to increase surface strength as measured by pick and resistance to linting are added at the size press.

Cellulosics. Carboxymethylcellulose (CMC) is the principal cellulose derivative used as a strength additive in papermaking. Like the anionic starches, CMC usually is used in alum-containing systems. When used in this way, CMC is about as effective a strength additive as the anionic acrylamide polymers. Carboxymethylcellulose also can be used with cationic, water-soluble, wet- and dry-strength additives. For example, when CMC is used with an aminopolyamide—epichlorohydrin wet-strength resin, increased wet and dry strength and sometimes a more uniform coating on Yankee dryers are achieved (see CELLULOSE ETHERS).

Acrylamide Polymers. In the early 1950s, anionic and cationic polyacrylamides were introduced as strength additives for paper (62). The advantage of these synthetic polymers over starches and natural gums is the ease with which molecular weight and degree of substitution of anionic and cationic groups can be varied. In general, the anionic acrylamide polymers, usually using acrylic acid as the comonomer to impart the negative charge, are used in alum-containing systems. In addition to improving the strength of paper, the acrylamide polymers often provide improved retention of fillers, improved rosin-sizing efficiency, faster drainage of the stock on the wire, and easier drying of the sheet.

The cationic acrylamide polymers may contain either tertiary amine or quaternary ammonium groups. Because of their positive charge, they are self-retaining on pulp fibers; therefore, they can be used effectively in nonalum systems.

Combinations of anionic and cationic resins are used. Some of the early systems involved the use of a cationic wet-strength resin with an anionic dry-strength additive to provide both increased wet and dry strength (63). Combinations of anionic and cationic dry-strength additives also are used to provide strength effects which cannot be achieved by using either polymer alone (64). The ratio of the two polymers must be optimized to achieve maximum performance (see ACRYLAMIDE POLYMERS).

Mechanism. The principal effect of a dry-strength additive is to improve the degree of bonding between fibers. Chemical strength additives provide, at least in part, the same effects as additional refining. In some cases, an additional effect is obtained and is comparable to that resulting from the use of stronger fibers. Of the total improvements in tensile and burst strengths, which are produced by locust bean gum, 60% are attributed to increased bond strength, 25% to improved sheet formation, and 15% to formation of new bonds (65). Thus, 75% of the dry-strength improvement is directly related

to improved bonding. Apparently the effect of a dry-strength additive is to provide extra adhesive for fiber bonding. This probably results from increasing the degree of the bonded area at each fiber—fiber crossover point as well as the frequency of bonding per unit fiber length. For example, the treatment of a paper sheet with a starch solution increases both the optical contact area per crossover region and the frequency of optical contact areas along the fiber. Further refining produces both of these effects. Thus, the addition of a dry-strength additive provides additional water-soluble polymeric adhesive to help bond the individual fibers.

Wet-Strength Additives. The dry strength of paper can be augmented by natural and synthetic polymers whose hydrogen-, ionic-, or covalent-bonding capabilities enhance the hydrogen bonding that already exists between cellulose fibers. In the presence of water, the cellulose hydrogen bonds are disrupted, and the paper does not have an appreciable amount of wet strength, ie, strength in the presence of water. With few exceptions, wet-strength additives (66) are capable of covalently bonding in order to preserve paper strength in the presence of water. Tissue and toweling, linerboard, medium, carrierboard, and bleached carton are some of the principal grades that require an amount of wet strength to be functional.

Urea—Formaldehyde and Melamine—Formaldehyde Resins. The first wet-strength resins were thermosetting urea—formaldehyde (UF) resins [9011-05-6] added at the wet end of the paper machine and cured in the drier section to produce paper with permanent wet strength. Anionic groups were incorporated in the UFs to enable them to be precipitated onto pulp with alum. Later, cationic UFs were produced by the addition of polyamines, such as diethylenetriamine [61472-52-4], that were self-retaining. Similarly, heating melamine (triamino-*s*-triazine) with three equivalents of formaldehyde yields trimethylolmelamine [1017-56-7] which, when treated with hydrochloric acid, forms a colloidal wet-strength resin. Both UF and melamine—formaldehyde (MF) resins are acid-curing, ie, their use requires mineral acid or, preferably, alum to catalyze the thermosetting reaction. On a weight basis, melamine resins can impart higher wet and dry strength than can UF resins. MF resins with lower residual (free) formaldehyde content, and hence lower subsequent emissions in the papermaking process, have been developed (67) (see AMINO RESINS AND PLASTICS).

Aminopolyamide—Epichlorohydrin Resins. If tissue and toweling is produced at a neutral-to-alkaline pH, the paper is softer and more absorbent which are desirable qualities for these grades. Thus, the aminopolyamide—epichlorohydrin resins, which cure most efficiently at neutral-to-alkaline pH, were readily accepted in these markets. In addition, these resins are effective creping aids in absorbent papers and allow for production of wet-strengthened paper without the negative aspects of acid-catalyzed degradation and embrittlement of the paper, and machine corrosion. As concerns grew about formaldehyde in the workplace, the use of aminopolyamide—epichlorohydrin resins surpassed UF and MF resins, eventually almost completely replacing them. The last grades to convert to polyamide resins were the high basis weight, unbleached grades which were difficult to repulp when wet-strengthened with polyamide resins. UF and MF grades repulp fairly readily; the repulping of polyamide-containing paper can be assisted by oxidants such as hypochlorite, which are, however, readily consumed by the lignin in unbleached fiber.

Polymeric Amine—Epichlorohydrin Resins. Various wet-strength resins have been produced utilizing still bottoms containing bishexamethylenetriamine [143-23-7], ie, by-products of hexamethylenediamine production. Like the aminopolyamide—epichlorohydrin resins, these resins are also neutral-to-alkaline curing, and their lower performance is offset by their lower cost. More recent polyamine—epichlorohydrin resins have been based on the use of poly(methyldiallylamine) [55553-13-4]. As these resins are provided in a storage-stable chlorohydrin form, they require alkaline activation to produce reactive epoxides, prior to use. They can be more efficient than the polyamide—epichlorohydrin resins, providing higher dry strength and dimensional stability on water immersion, although they can also be more difficult to repulp.

Aldehyde-Modified Resins. Polymers containing reactive aldehyde groups, as distinguished from those made with formaldehyde and containing methylol groups, are characterized by good absorbency, high dry-strength efficiency, and fugitive wet strength, which makes them more suitable for sanitary papers than packaging grades. The wet strength that is provided by aldehyde-modified resins decreases faster and further, to below one-third its initial value after prolonged soaking, than that imparted by amino polymer—epichlorohydrin resins. Commercial versions of these temporary wet-strength resins include cationic polyacrylamide-glyoxal resins and cationic blocked aldehyde starch dispersions (31,32). These resins are most efficient at lower pH, especially in the presence of alum, but can develop wet strength in paper as high as pH 6.0–6.5.

Mechanism of Wet-Strength Development. There are two predominant theories that attempt to explain the mechanism of wet-strength development in paper. The protection theory proposes that the wet-strength resin forms a restraining network, by cross-linking either with itself or bonding with the cellulose. Thus, the network protects a fraction of the hydrogen bonding in the dry sheet by limiting the swelling of the cellulose and hemicelluloses in the presence of water. Alternately, the reinforcement theory proposes that the new covalent bonds formed by the wet-strength resin are the ones that remain unbroken by water.

In support of the protection theory are observations that, in the curing reaction, urea—formaldehyde resins have approximately the same activation energy on a variety of substrates, including glass fiber and cellulosic pulps (68). Also, the strength which is imparted by a given amount of aminopolyamide—epichlorohydrin resin to handsheets of different pulps tends to be a relatively constant percentage of the dry strength, even though the absolute values of dry and wet strength vary widely. The new bond theory is supported by observations that wet and dry strength increase by approximately equal increments using acid-curing urea—formaldehyde and melamine—formaldehyde resins or using alkaline-curing aminopolyamide—epichlorohydrin resins (69–71).

Whatever its mechanism, the development of wet strength usually requires formation of a covalently bonded network of resin-to-resin or resin-to-cellulose bonds, or both. To optimize the performance of a cationic wet-strength additive, especially the aminopolyamide—epichlorohydrin type, the chemical should be added to the long fiber furnish, to minimize contact and adsorption on cellulose fines. Resin adsorbed on fines contributes much less to the paper's ultimate wet and dry strength than resin adsorbed on the long fiber portion (72). In urea—formaldehyde or melamine—formaldehyde resins, alum or acid catalyzes insolubilization of resin through methylene ether cross-links with expulsion of water. In alkaline-curing resins, epoxy or 3-hydroxyazetidinium groups cross-link with amine groups on the resin (73) or with cellulose hydroxyls and carboxyls (74). In both kinds of resins, the cross-links are difficult to hydrolyze; therefore, they impart wet strength which survives prolonged soaking. The aldehyde resins, eg, blocked aldehyde—starch and polyacrylamide—glyoxal, probably react with cellulose hydroxyls to form hemiacetal bonds which, though covalent, are hydrolyzed rapidly, resulting in the fugitive wet strength discussed previously.

Environmental Concerns. Concern about formaldehyde in the workplace has caused a permanent shift from UF and MF resins to polyamide—epichlorohydrin resins. UF and MF products typically contained 0.5–3.0% free formaldehyde in order to maintain stability and performance. Concern has been raised about whether a workplace hazard might occur owing to the presence of free epichlorohydrin in those neutral-to-alkaline resins produced with it. However, the residual epichlorohydrin in these resins has been well below the current reporting requirements for these materials (SARA regulations require products containing >1000 ppm (0.1%) of epichlorohydrin to be labeled), and often below detection limits (at 1 ppm by gas chromatography mass spectrometry) (75). Concern has been raised about the presence of 1,3-dichloro-2-propanol, a hydrolysis product of epichlorohydrin, in products made with epichlorohydrin. Though not currently regulated, products have been designed to have minimal levels of 1,3-dichloropropanol (<1000 ppm) and performance levels similar to those of polyamide—epichlorohydrin resins (76–79). Obviously, the same concern could be raised about any product manufactured with epichlorohydrin, including many retention-aid polymers. However, in monitoring of workplace air, the presence of epichlorohydrin or 1,3-dichloro-2-propanol is usually not detected except in the headspace of improperly vented storage tanks (75). For air monitoring, the current detection limit is about 0.5 ppm. With formaldehyde-containing strength additives, significant levels of formaldehyde can be detected in the workplace, at both the paper mill and at subsequent converting operations when the paper is rewet and dried.

The proposed cluster rules would limit the discharge of adsorbable organic halide (AOX) from an integrated bleached pulp paper mill (80). Though most of the AOX in mill effluent originates in the bleaching process (and some is naturally occurring in the trees prior to pulping (81)), concern has been raised about the organic chloride-containing components present in products manufactured with epichlorohydrin. Products that have been designed to contain reduced levels of 1,3-dichloro-2-propanol also have reduced levels of AOX (76–79). Additionally, the organic chloride-containing components of paper additives are most often small, highly water-soluble molecules that degrade in mill effluent treatment systems and also would not tend to bioaccumulate in fatty tissues (82).

Fillers. Opacity must be high enough in paper with print or writing on both sides of the sheet to prevent images from showing through from the back side. Newsprint typically has had sufficient opacity because the many small groundwood fibers provide much greater surface area from which light can scatter than the larger fibers in typical chemical pulps. In more permanent, higher quality papers, the chemically pulped, bleached fibers are not high in scattering power and do not produce high opacity paper, particularly when the basis weight of the paper is low, eg, <65 g m². Opacity is increased by

incorporating particulate materials or fillers into the sheet in order to increase the scattering of light which passes through the sheet (see [FILLERS](#)).

Though functionally and chemically similar, fillers and pigments are distinguished from one another in that fillers are added at the wet end of the paper machine, and serve to fill the sheet; pigments are added at the size press and serve to alter the surface of the sheet. The most common fillers are mineral pigments, eg, clay, titanium dioxide [[13463-67-7](#)], calcium carbonate, silica [[7631-86-9](#)], hydrated alumina [[21645-51-2](#)], and talc [[14807-96-6](#)]. Kaolin clay [[1332-58-7](#)] is the least expensive and most widely used filler pigment in the United States (83). Methods have been devised to render almost any filler cationic, thus enhancing its retention on the fiber (84). Calcium carbonate has been limited to use in papermaking systems where the pH is neutral or slightly alkaline, because it dissolves in low pH systems. Until recently, most of the calcium carbonate that was used as filler for paper in the United States was produced by a precipitation process (PCC). Newer developments in surface-modified PCC may allow for its use in more acidic environments (85). The use of naturally occurring ground calcium carbonates (GCC) as fillers is increasing in the United States because PCC has been shown to have a more negative impact on internal sizing than does GCC (86,87). In Europe, natural calcium carbonates are widely used because they are readily available in suitable, small-particle size and are less expensive than clay.

Overall, the use of these fillers in paper has been increasing because they are less expensive than pulp fiber. Also, as the trend toward lower basis weight continues, the use of filler increases to compensate for opacity losses. This has been especially true in newsprint, where demands on the sheet for print performance coupled with reduced basis weight has led to a significant use of filler (88,89). As the bulk of the sheet is enhanced by filler at the expense of fiber, the loss in sheet strength has to be compensated for, mechanically or chemically. Additionally, the higher surface area of fillers can consume papermaking additives, eg, sizing and strength additives whose destination was to be the fiber, necessitating an increase in the use of these additives to achieve the targeted paper properties (87). Cationically modified fillers do not cause quite as great a deterioration of sizing chemistry (90) as do the anionic precursors.

The previously described mineral fillers function as opacifiers primarily by increasing the amount of surface area in the paper sheet and thus increasing the scattering of light. All of these fillers are characterized by refractive indexes that are similar to that of the cellulose fiber, ie, 1.53. Titanium dioxide not only increases surface area but scatters light within the particle because of its high refractive index, ie, 2.5–2.7. Unlike the other mineral pigments, titanium dioxide is considerably more expensive than the pulp fiber in the paper. Therefore, it is used in grades of paper that require high opacity and high brightness (91).

In recent years, synthetic polymeric pigments have been promoted as fillers for paper. Pigments that are based on polystyrene [[9003-53-6](#)]/latexes and on highly cross-linked urea–formaldehyde resins have been evaluated for this application. These synthetic pigments are less dense than mineral fillers and could be used to produce lightweight grades of paper, but their use has been limited in the United States.

Functional Surface Treatments

Although many functional chemicals can be added to the wet end of the paper machine, some grades of paper require special properties that cannot be provided by the low levels of wet-end additives that are retained in the interior of the sheet of paper. Examples are high quality printing and writing paper, which require high levels of ink holdout and surface strength, and ink jet printing paper, which requires high levels of ink holdout, resistance to wicking, and resistance to color-to-color bleed. To achieve the properties required for these grades of paper, it is necessary to apply special chemicals to the surface of the preformed paper web.

Processes. The most common method for the application of chemicals to the surface of a paper web is by a size press. In the size press, dry paper, which usually is sized to prevent excess water and chemical penetration, is passed through a flooded nip or pond, and a solution or dispersion of the functional chemical contacts both sides of the paper. Excess liquid is squeezed out in a press and the paper is redried.

The gate-roll size press is used for the application of high solids, high viscosity compositions to the surface of the sheet (92). The material to be applied is transferred over a series of rolls, and a thin film from the applicator roll is applied to the sheet.

At high speeds, the pond of the flooded nip size press becomes turbulent. Roll maintenance is a problem with the gate-roll size press. To avoid these problems, the blade rod metering size press was developed. Short-dwell coater heads are used to apply a precisely controlled quantity of chemicals to the size press rolls. This quantity is controlled with either a metering blade or a metering rod. Blade or rod metering eliminates the pond, and does not increase the number of rolls required for surface chemical application.

Spray applications to the surface of the sheet have been useful, especially for application of creping aids and release agents in towel and tissue mills. The spray application of functional chemicals has not been used widely. Generally, the uniformity of application is more difficult to control when functional chemicals are sprayed than when they are applied by a size press. Functional chemicals also can be applied to heavier grades of paper or paperboard at the calendar stack.

Sizing. The most commonly used materials for surface sizing are starches and modified starches, including oxidized [[65996-62-5](#)], enzyme-converted [[65996-64-7](#)], hydroxyethylated [[9005-27-0](#)], and cationic starches. They are used not only for sizing, but also to improve strength, especially surface strength, and to impart smoothness. Starches may be applied to the finished sheet by any of the previously discussed methods. Often starch is used with other surface sizing agents, such as synthetic polymeric sizing agents. These combinations permit improved sizing against liquid penetrants, increased surface strength, and better finish.

Synthetic polymeric sizing agents contain hydrophobic elements and water-soluble functionalities. The two most popular classes of synthetic polymeric sizing agents are styrene–maleic anhydride copolymers (93) and polyurethane dispersions. Styrene–maleic anhydride copolymers (SMAs) use styrene as the hydrophobic element and the hydrolyzed maleic anhydride ring as the water-soluble functionality. These copolymers are extremely popular because they improve a large number of paper properties (94). These film formers improve sizing, improve printability, improve surface strength, increase paper surface coefficient of friction, and decrease paper porosity or permeability to air flow.

Polyurethane dispersions are used in specialty applications, where high levels of sizing are needed. Wax emulsions and wax–rosin emulsions also are used by themselves as surface-applied sizing agents to produce very high resistance to liquid penetration in paper and paperboard. Other products that are used as surface sizes include CMC and poly(vinyl alcohol), which provide oil- and grease-repellent coatings, improve paper strength, and decrease paper porosity.

Alkylketene–dimer emulsion sizes can be applied to the surface of paper and provide very efficient sizing. Used by themselves, these sizing agents can provide a slippery surface; consequently, they often are used in conjunction with starch or some filler which reduces slipperiness without detracting from sizing efficiency.

Fluorochemical emulsion sizing agents can be applied to the surface of paper or paperboard to provide good oil and grease repellency. If they are used with other sizing agents, eg, alkylketene dimer emulsions, both oil and grease repellency and water repellency are obtained. Fluorochemical surface treatments are used for pet-food bag paper, labels, coupons, cookie bags, candy wrappers, snack food bags, reprographic papers, and meat, fish, and poultry wrap.

Application of Dry-Strength Additives. The various water-soluble natural and synthetic polymers which are used for strength enhancement by internal addition can be applied to paper surfaces. This type of application usually is indicated when surface strength properties are more important than increased internal strength. Starches and modified starches, especially cationic starches, are used in large quantities to improve the surface strength of paper; they also improve the printing quality of the paper as a result of increased surface strength and reduced linting. The natural gums, eg, guar, locust bean, and tamarind, also can be applied to the surface of the paper to enhance strength. Likewise, various derivatives of these natural gums can be used. The acrylamide polymers normally are not used for surface applications, since they can be retained effectively during the papermaking operation. However, they can be used where primarily surface strength improvement is desired. All of these applications can be made by any of the techniques previously described.

Application of Wet-Strength Resins. Wet-strength resins seldom are applied to the surface of paper for enhancing wet strength because the commercially available, cationic wet-strength resins are retained so effectively internally. However, wet-strength resins are applied frequently to the surface of towels and tissues as creping aids.

Creping. The products that are used as internally added creping aids can be applied to the surface of paper, usually by spraying an aqueous solution or emulsion in front of the Yankee dryer. One of the most commonly used resins for this purpose is an aminopolyamide–epichlorohydrin resin. Such resins yield coatings on the dryer with the required degree of adhesion for optimum creping. If the internally added wet-strength resin yields excessive adhesion to the dryer, a release agent, eg, a silicone–oil emulsion, can be sprayed on the sheet. Application of creping aids to the sheet in front of the dryer permits rapid manipulation of the degree of adhesion by varying the ratio of adhesive to release agent.

Curl Control. Many grades of paper tend to curl, especially as humidity varies, because of the stresses produced during the drying process. This is especially troublesome when only one side of the paper receives a surface treatment. Judicious application of water to the opposite side of the dry sheet followed by redrying may correct the curling. Water may be applied by surface application at the size press, water box, or calender stack or it may be sprayed on. Small amounts of water can be applied to the paper surface as a foam with excellent results.

Pigmented Coatings. Conventional paper has a surface that is not well suited to high speed printing processes, principally because the surface is rough; thus, contact of the surface with printing elements is poor. Furthermore, in quality printing, a high level of gloss in both the printed areas and in the substrate often is desirable. High levels of gloss cannot be attained with the rough surface of conventional paper. For these reasons, a large volume of paper is coated in order to improve printability. The main function of the coating is to provide a smooth surface for printing. Other properties that are important to the coating are receptivity to inks and sufficient surface strength to withstand the forces of the printing process on the coated paper.

Paper coatings are applied as coating colors, which are aqueous slurries containing 35–65 wt % solids. There are three main components of the solids; pigments, binders, and minor additives. The pigment is the primary component of a paper coating and consists of small, white, particulate material. Pigments usually are minerals, eg, clay, calcium carbonate, or titanium dioxide. The packed pigment particles fill pitted areas of the rough paper surface, thereby providing a suitable surface for printing. Binders are the resins or polymers that function as the glue that binds the pigment particles to each other and to the paper substrate. The level of binder is low in a paper coating, typically 5–30 parts by weight per 100 parts of pigment. This low level of binder distinguishes paper coatings from paints, which are pigment-filled polymer films. Minor additives are used to modify the properties of the coating color, primarily before and during the coating operation.

Pigments. The pigments which are used in paper coatings are similar to the materials that are used as fillers (83,88). Kaolin clay is the largest-volume pigment for this application. The clay that is used in coatings is of better quality and is slightly more expensive than that used as a filler. Coating clays are higher in brightness and have a smaller particle-size distribution than the filler clays. Structured kaolin pigments (95,96), in which the base clay is aggregated or flocculated rather than calcined, provide enhanced opacity and oil absorption.

Finely ground calcium carbonate is second in volume to kaolin clay as a paper-coating pigment. Titanium dioxide is used in coated grades of paper requiring high levels of brightness and opacity. Preblending calcium carbonate with titanium dioxide produces a pigment that scatters light almost as well as titanium dioxide alone, but at a lower cost (97). Plastic pigments that are based on polystyrene are used in combination with mineral pigments to improve the gloss of coated paper. The use of hollow acrylate polymer latex particles in coating formulations has shown promise for the gloss-enhancement of paper and paperboard (98).

Satin white [12004-14-7] is another paper-coating pigment and is prepared by the reaction of calcium hydroxide with alum. In combination with other mineral pigments, satin white produces coatings with improved ink receptivity, gloss, smoothness, and brightness. Satin white usually is used as <25 wt% of the pigment system, because it causes rheological problems at higher levels. Most satin white is produced by paper companies for their own use.

Binders. Paper-coating binders are either polymers derived from natural sources or synthetic polymers. The largest volume, naturally derived binder is starch (qv) (99). Starch is provided in derivatized form or unmodified form; pearl corn starch is used predominantly for the latter. Unmodified starch is solubilized by thermal conversion or enzyme conversion. The most common derivatized products are acetylated [9045-28-7], oxidized, and hydroxyethylated starches.

The other main natural binder is protein that is prepared by extraction from soy meal (100). Casein [9000-71-0], once a large-volume paper-coating binder, has markedly declined in use because of its high price and susceptibility to microbial attack.

Almost all synthetic binders are prepared by an emulsion polymerization process and are supplied as latexes which consist of 48–52 wt % polymer dispersed in water (101). The largest-volume binder is styrene–butadiene copolymer [9003-55-8] (SBR) latex. Most SBR latexes are carboxylated, ie, they contain copolymerized acidic monomers. Other latex binders are based on poly(vinyl acetate) [9003-20-7] and on polymers of acrylate esters. Poly(vinyl alcohol) is a water-soluble, synthetic binder which is prepared by the hydrolysis of poly(vinyl acetate) (see LATEX TECHNOLOGY; VINYL POLYMERS).

Other. A large variety of additives are used in paper-coating colors primarily to modify the physical properties of the colors (102). At high solids concentrations in water, mineral pigment particles tend to associate and form viscous pastes. Dispersants (qv) are used to prevent this and to provide low viscosity slurries. Common dispersants include polyphosphates and sodium polyacrylate [9003-04-7]. Various water-soluble polymers are added to coating colors and act as water-retention agents and as rheology modifiers.

Hydrophilic polymers function as water-retention aids by preventing premature dewatering of the coating color after it has been applied to the paper but before the paper has been dried. Water-soluble polymers, eg, CMC, hydroxyethylcellulose [9004-62-0], guar gum and derivatives, and sodium alginate [9005-38-3], improve the rheological properties of coating colors and help keep the colors on the surface of the paper rather than striking into the sheet. Lubricants are added to coating colors to improve the lubricity of the wet coating color and to improve the properties of the dried coating. In particular, lubricants prevent sticking of the dry coatings to surfaces of calenders. Common lubricants include calcium stearate [1592-23-0], fatty acid esters, sulfonated oils, and wax emulsions.

Synthetic Fibers

A variety of wet-laid felts and nonwoven fabrics are produced on Fourdrinier-type paper machines (103,104) (see NONWOVEN FABRICS). Noncellulosic materials may be included as part or all of the fiber furnish; latexes, water-soluble polymers, or other adhesives are used as bonding agents. Synthetic fibers can make paper highly resistant to wetting; chemical attack; mechanical wear, eg, folding; weathering; and biological degradation. Synthetic fiber-containing papers are used as backings for carpets and vinyl floor coverings, industrial filters, disposable bed linens and hospital garments, heavy-duty wiping materials, tea bags, tissues, labels, and embossable wall papers.

Water-laid cellulose papers contain strong interfiber bonds which result from overlap, interpenetration, and consolidation of swollen fiber surfaces. The main synthetic fibers swell little, if at all, in water, and require special bonding techniques whether they are water-laid or air-laid. Bonding techniques include swelling of fiber crossover points by organic solvents or by salt solutions, which become concentrated during drying, and controlled melt bonding. Bonding with fibrils depends on fusion and entanglement of small, branched, fibrillated fibers (105,106). Nylon and polyester fibers improve wet strength, tear strength, and fold resistance of papers (107).

Heavy-duty wiping materials and some disposable garments are made from nonwoven cellulose mats which are reinforced with a synthetic fiber, eg, nylon scrim network; the spacing of the synthetic threads is 0.5–1.0 cm (108).

Asbestos [1332-21-4] fiber traditionally is used with a latex binder in backings for roll-vinyl floor coverings, but the long-term health hazards which asbestos (qv) presents have initiated a search for substitutes, eg, glass wool, rock wool, polyolefin fibers, and cellulose. Polyolefin fibers have been investigated extensively as substitutes for cellulose and asbestos fibers in papermaking because of the formers' good optical properties, low density, thermoplasticity, inertness, and machine productivity. Because they float on water and have hydrophobic surfaces, polyolefin fibers require chemical or mechanical pretreatments so that they can be dispersed in water. Although polyolefin fibers cost more than cellulose, they are cost-effective because their low density and high surface area:weight ratio provides comparable print quality and opacity at lower sheet weights than do all-cellulose papers. Polyolefin–cellulose composites also are used in nonasbestos flooring felts, wallpapers, filter media, labels, embossable papers, and other nonwoven fabrics that are made on paper machines. Use of synthetic fibers in paper has been reviewed (103,104).

BIBLIOGRAPHY

"Papermaking Additives" in *ECT* 3rd ed., Vol. 16, pp. 803–825, by S. T. Putnam, H. H. Espy, and G. G. Spence, Hercules Inc.

1. H. K. Berry, *Papermaker* (2), 35–37 (1994).
2. B. A. Egan and R. Dickinson, *Pulp Paper*, **64**(6), 110–113 (1990).
3. *Title 21 Code of Federal Regulations*, Subchapter B, Part 176, Sections 170, 180, U.S. Government Printing Office, Washington, D.C., Apr. 1, 1992.
4. D. E. Boardman, *Tappi* **76**(12), 148–152 (1993).
5. D. Guest, *Paper Technol.* **31**(6), 40–43 (1990).
6. *Pulp and Paper 1993 Buyers Guide Issue*, Miller Freeman, Inc., San Francisco, Calif., 1993, pp. 93–111.
7. R. M. Husband, ed., *Survey of Paper Additives*, 2nd ed., H&H Consulting Group, Trumbull, Conn., 1989.
8. L. H. Allen and I. M. Yaraskavitch, *Tappi* **74**(7), 79–84 (1991).
9. J. E. Lukins, *Pulp Paper* **67**(4), 69–71 (1993).
10. P. Arvela, J. W. Swanson, and R. A. Stratton, *Tappi* **58**(11), 86 (1975).
11. W. E. Frankle and J. L. Sheridan, *Tappi* **59**(2), 84 (1976).
12. K. W. Britt, *Tappi* **56**(3), 83 (1973).
13. K. W. Britt, *Tappi* **56**(10), 46 (1973).
14. **U.S. Pat. 5,185,062** (Feb. 9, 1993), A. J. Begala (to Nalco Chemical Co.).
15. **U.S. Pat. 5,266,164** (Nov. 30, 1993), R. W. Novak and T. C. Fallon (to Nalco Chemical Co.).
16. L. Wagberg, X. P. Zhao, I. Fineman, and F. N. Li, *Tappi* **73**(4), 177–182 (1990).
17. B. Carre, *Nordic Pulp Paper Res. J.* **8**(1), 21–26 (Apr. 1993).
18. R. I. S. Gill, "Developments in Retention Aid Technology," *Paper Technol.*, 6–11 (Aug. 1991).
19. P. Kubelka and F. Munk, *Z. Tech. Phys.* **12**, 593 (1931).
20. P. Kubelka, *J. Opt. Soc. Am.* **39**(5), 448 (1948).
21. J. A. Van den Akker, *Tappi* **32**(11), 498 (1949).
22. F. A. Steele, *Paper Trade J.* **100**(12), 37 (1935).
23. K. W. Britt and J. E. Unbehend, *Tappi* **59**(2), 67 (1976).
24. P. F. W. Lee and T. Lindstrom, *Nordic Pulp Paper Res. J.* **4**(2), 61–70 (July 1989).
25. **U.S. Pat. 5,185,061** (Feb. 9, 1993), P. Lowry and D. Farrar (to Allied Colloids Ltd.).
26. K. W. Britt and J. E. Unbehend, *Tappi Papermakers Conference Proceedings 1980*, Tappi Press, Atlanta, Ga., 1980, pp. 5.
27. L. F. Twoomey, *Tappi* **70**(9), 87–90 (1987).
28. R. H. Voss, C. E. Luthe, B. I. Fleming, R. M. Berry and L. H. Allen, "Some New Insights into the Origins of Dioxins Formed During Chemical Pulp Bleaching," paper presented at the *CPPA Environment Conference 1988, Vancouver, B.C., Canada*, Oct. 25–26, 1988.
29. *Questions and Answers: Water-Based Antifoams*, Nalco Chemical Co., Oak Brook, Ill., 1983.
30. M. Laleg and I. I. Pikulik, *Nordic Pulp Paper Res. J.* **6**(3), 99–109 (Oct. 1991).
31. L. R. Peel, *Paper Technol.* 20–23 (Dec. 1989).
32. M. Laleg and I. I. Pikulik, *J. Pulp Paper Sci.* **17**(6), J206–J216 (1991).
33. S. D. Goldstein, *Tappi* **75**(7), 85–92 (1992).
34. L. H. Allen and C. L. Lapointe, *Nordic Pulp Paper Res. J.* **4**(2), 94–98 (July 1989).
35. T. Hassler, *Tappi* **71**(6), 195–201 (1988).
36. D. S. Rende, *Pima* **76**(2), 32–33 (1994).
37. **U.S. Pat. 5,139,616** (Aug. 18, 1992), T-F. Ling (to Betz PaperChem, Inc.).
38. J. A. Koncel, *Amer. Papermaker* **54**(2), 50–51 (1991).
39. J. A. Glazer, *Tappi* **74**(7), 72–74 (1991).
40. A. S. Nigrelli, *Pulp Paper* **64**(11), 68–72 (1990).
41. J. P. Oliver, *Tappi* **63**(12), 91–95 (1980).
42. R. W. Davison, *Tappi* **58**(3), 48 (1975).
43. E. W. Washburn, *Phys. Rev.* **17**, 273 (1921).
44. J. Marton, *Nordic Pulp Paper Res. J.* **4**(2), 77–80 (July 1989).
45. J. Marton and F. L. Kurlle, *J. Pulp Paper Sci.* **13**(1), J5–J9 (June 1987).
46. L. J. Stryker, B. D. Thomas, Jr., and E. Matijevic, *J. Colloid Interface Sci.* **43**(2), 319–329 (May 1973).
47. J. Marton and F. L. Kurlle, *Canad. J. Pulp Paper Sci.* **13**(1) (1987).
48. B. Wortley, *Pulp Paper*, **64**(11), 131–133 (1990).
49. M. C. Peck and M. A. Markillie, *Tappi Papermakers Conference Proceedings 1994*, Tappi Press, Atlanta, Ga., 1994, p. 165–168.
50. J. Marton, *Tappi* **73**(11), 139–143 (1990).
51. B. W. Crouse and D. G. Wimer, *Tappi* **74**(7), 152–159 (1991).
52. W. J. Andounian, *Pima* **75**(7), 48–49 (1993).
53. G. B. Kelly, Jr., *Alkaline Paper Advoc.* **2**(1), 9–10 (Apr. 1989).
54. J. Marton and D. D. Jarrell, *Sizing Short Course 1987*, Tappi Press, Atlanta, Ga., pp. 53–62.
55. A. R. Colasurdo and I. Thorn, *Tappi* **75**(9), 143–149 (1992).
56. R. W. Novak and D. S. Rende, *Tappi* **76**(8), 117–120 (1993).
57. J. Marton, *Tappi* **74**(8), 187–191 (1991).
58. *TAPPI Test Methods, 1992–1993*, Tappi Press, Atlanta, Ga., 1992.
59. C. E. Faley, *Tappi Papermakers Conference Proceedings 1986*, Tappi Press, Atlanta, Ga., pp. 147–150.
60. F. M. K. Werdouschegg, in W. F. Reynolds, ed., *Dry Strength Additives*, Tappi Press, Atlanta, Ga., 1980, pp. 67–93.
61. M. Laleg and I. I. Pikulik, *J. Pulp Paper Sci.* **19**(6), J248–J255 (Nov. 1993).
62. W. F. Reynolds, in Ref. 60, pp. 125–148.
63. **U.S. Pat. 3,049,469** (Aug. 14, 1962), R. W. Davison (to Hercules Inc.).
64. **U.S. Pat. 3,332,834** (July 25, 1967), W. F. Reynolds (to American Cyanamid Co.).
65. H. J. Leech, *Tappi* **37**(8), 343 (1967).
66. *Tappi Wet and Dry Strength Short Course 1988*, Tappi Press, Atlanta, Ga., 1988.
67. W. Kamutzki, *Polym. Paint Colour J. (Suppl.)*, 108–114 (1989).
68. H. L. Marder, S. E. Church, and V. Stannett, *Tappi* **40**(10), 829 (1957).
69. A. Jurecic, C. M. Hou, K. Sarkanen, C. P. Donofrio, and V. Stannett, *Tappi* **43**(10), 861 (1960).
70. D. J. Salley and A. F. Blockman, *Tech. Assoc. Pap.* **30**, 223 (1947); *Pap. Trade J.* **125**(1), 35 (1947).
71. A. Jurecic, T. Lindh, S. E. Church, and V. Stannett, *Tappi* **41**(9), 465 (1958).
72. R. A. Stratton, *Nordic Pulp Paper Res. J.* **4**(2), 104–112 (July 1989).
73. D. I. Devore and S. A. Fischer, *Tappi* **76**(8), 121–128 (1993).

74. H. H. Espy and T. W. Rave, *Tappi* **71**(5), 133–137 (1988).

75. M. A. Dulany, technical data, Georgia-Pacific Paper Chemicals Laboratory, Decatur, Ga., 1992–1994.

76. D. I. Devore, N. S. Clungeon, and S. A. Fischer, *Tappi* **74**(12), 135–141 (1991).

77. **U.S. Pat. 5,019,606** (May 28, 1991), M. Marten and W. Kamutzki (to Hoechst, AG).

78. **U.S. Pat. 5,189,142** (Feb. 23, 1993), D. I. Devore and S. A. Fischer (to Henkel Corp.).

79. **U.S. Pat. 5,239,047** (Aug. 24, 1993), D. I. Devore and S. A. Fischer (to Henkel Corp.).

80. A. Rooks, *Pima* **75**(11), 34–40 (1993).

81. O. Dahlman and co-workers, *Environ. Sci. Technol.* **27**(8), 1616–1620 (1993).

82. G. R. Craig, P. L. Orr, J. L. Robertson and W. M. Vrooman, *Pulp Paper Can.* **91**(9), 39–45 (1990).

83. R. W. Hagemeyer, ed., *Pigments for Paper*, Tappi Press, Atlanta, Ga., 1984.

84. **U.S. Pat. 5,169,441** (Dec. 8, 1992), R. V. Lauzon (to Hercules Inc.).

85. M. Tokarz, R. Sikkar, and S. Jaras, *Nordic Pulp Paper Res. J.* **5**(2), 88–92 (July 1990).

86. M. C. Peck, technical data, Georgia-Pacific Paper Chemicals Laboratory, Decatur, Ga., 1991–1993.

87. J. Marton, in *Papermaking Raw Materials*, Vol. 2, Mechanical Engineering Publications Ltd., London, pp. 785–813 (1985).

88. A. Rook, *Pima* **75**(7), 39–45 (1993).

89. C. A. Skrocki and C. D. Richardson, *Pulp Paper*, **64**(2), 65–69 (1990).

90. **U.S. Pat. 5,147,507** (Sept. 15, 1992), R. A. Gill (to Pfizer, Inc.).

91. B. Siffert and J-M. Metzger, *Coll. Surf.* **53**, 79–99 (1991).

92. R. J. Alheid, *Tappi Papermakers Conference Proceedings 1978*, Atlanta, Ga., p. 83.

93. G. L. Batten, Jr., *Tappi Papermakers Conference Proceedings 1992*, Atlanta, Ga., pp. 159–168.

94. T. W. Tompkins, in Ref. 93, pp. 169–174.

95. R. W. Dahlquist, R. E. Hardy, and L. J. Welch, *Pulp Paper* **63**(11), 90–93 (1989).

96. T. L. Kellogg and R. E. Hardy, *Pulp Paper*, **67**(5), 71–74 (1993).

97. W. E. Scott, *Am. Papermaker* **50**(5), 12–14 (1987).

98. **U.S. Pat. 5,157,084** (Oct. 20, 1992), D. I. Lee, M. R. Mulders, D. J. Nicholson, and A. N. Leadbetter (to The Dow Chemical Co.).

99. *Tappi Monogr. Ser.* **17** (1957).

100. *Tappi Monogr. Ser.* **36** (1975).

101. *Tappi Monogr. Ser.* **37** (1975).

102. C. G. Landes and L. Knoll, eds., *Paper Coatings Additives*, Tappi Press, Atlanta, Ga., 1978.

103. O. A. Battista, ed., *Synthetic Fibers in Papermaking*, Wiley-Interscience, New York, 1964.

104. R. W. Moncrieff, *Man-Made Fibres*, Halstead Press, a division of John Wiley & Sons, Inc., New York, 1975.

105. R. A. A. Hentschel, *Tappi* **42**(12), 979 (1959).

106. L. A. Auspos and E. B. Winn, *Tappi* **45**(9), 741 (1962).

107. G. L. McLeod, *Tappi* **41**(8), 430 (1958).

108. **U.S. Pat. 3,616,133** (Oct. 26, 1971), G. D. Thomas (to Kimberly-Clark Corp.).

Margaret A. Dulany
George L. Batten, Jr.
Michael C. Peck
Charles E. Farley
Georgia-Pacific Resins, Inc.

PARAFFIN WAX.

See WAXES.

PARALDEHYDE.

See ACETALDEHYDE.

PARASITIC INFECTIONS, CHEMOTHERAPY.

See ANTIPARASITIC AGENTS; ANTHELMINTICS; ANTIMYCOTICS.

PARATHION.

See INSECT CONTROL TECHNOLOGY; PESTICIDES; PHOSPHORUS COMPOUNDS.

PARTICLE SIZE MEASUREMENT.

See SIZE MEASUREMENT OF PARTICLES.

PARTING AGENTS.

See RELEASE AGENTS.

PATENTS AND TRADE SECRETS

Practice and Management

This article provides a basic, step-by-step approach to problem solving in the practice and management of patents and trade secrets. The significance of aggressive patent and trade secret protection to the economic well-being of a business or organization should not be underestimated. Without patents and trade secrets, the marketplace is reduced to competition on the basis of price, which may be very difficult.

Patents and trade secrets are protected by securing rights to ideas and the application of ideas that have commercial worth. The grant of rights in patents and trade secrets is based on an appreciation of development, advancement, and invention that will stimulate innovation by advancing technology. Patents and trade secrets are two distinct mechanisms for protecting invention vis-a-vis the application of ideas. Both are supported by the policies and laws of the United States.

Compiling a portfolio of patents provides an organization with an offensive weapon with which to protect and ensure profitability. Competitors who desire to enter a product market or maintain a meaningful interest within it must engage the owner of the leading technology in that market. If the leading technology is protected by patents, the owner of this technology has an excellent tool with which to ensure their profitability. A patentee may in turn realize a return on the time and energy invested in obtaining protection by securing a principal interest in the market, royalties from competitors, or even damages from those who choose to ignore the rights flowing from the patent grant (see LICENSING).

A patent also serves a defensive function. It provides the patentee with a partial or total shield that prevents others from patenting inventions which would ultimately restrict the patentee's commercial activity in the marketplace.

Trade secret rights are based on the complete absence of disclosure of the invention to anyone other than the owner. Oftentimes ideas, developments, and advances that are the subject of trade secret protection are those which may not be patentable, for any of a number of reasons. These reasons can include the nature and subject matter of the advance or development, as well as the commercial value of the advance or development. In any instance, an individual, business, or corporation is well-advised to consider all possible means of protection when reviewing an advance, development, or invention.

Some factors to consider when evaluating patent and trade secret protection include (1) the form and content of the technological advance, idea, development, or application; (2) the desired term of protection; (3) the potential for the technological advance, idea, development, or application to be the subject of a commercial product; (4) work done previously; (5) events which have publicized or publicly disclosed the technological advance, idea, development, or application; and (6) factors that may be critical to keeping the technological advance, idea, development, or application confidential, and what events may necessitate disclosure.

Harmonization

During 1995 the U.S. Patent laws changed to comply with certain international conventions. Two principal conventions, the General Agreement on Tariffs and Trade (GATT) as well as the North American Free Trade Agreement (NAFTA) have effected a change in the term of patents issued from the United States Patent and Trademark Office (U.S. PTO); a change in type of patents that may be filed in the U.S. PTO; and prospective changes that will

internationalize U.S. patent law.

By the turn of the century the U.S. PTO may be operating under a system that includes (1) publication of patent applications; (2) opposition of allowed applications for purposes of testing validity; (3) the dawn of first-to-file priority examination; and (4) the end of the antiquated test of inventorship called "interference practice." Legislation implementing many of these changes is pending before the U.S. Congress.

The Origin of Patent Rights

A patent is an affirmative right granted by the U.S. Federal Government. The affirmative right is represented by a published written document, referred to as a patent, which provides a full and complete description of the invention. The affirmative rights which stem from the issuance of a patent allow the owner of that patent to prevent other parties from making, using, or selling in the United States what the patent covers. The coverage of a patent is the actual property of the patent owner and is defined by the patent claims, which are like the legal description of real estate in a deed. Interpretation of the patent claims involves answering complex legal questions and is dependent on, among other things, the written description in the patent.

The printed published document which represents the patent rights granted by the Federal Government can be a complex literary work. There are specific and rigid legal requirements for the description, disclosure, and definiteness which support these affirmative rights and enable enforcement of those rights by the inventor or owner of the patent. The basis for this full and complete disclosure of the invention in the patent is clearly articulated in the U.S. Constitution.

A patent is intended to further the development of science and technology by providing a published record of technological developments for all to read, consider, and discuss. At the same time, a patent provides a delineation or definition of the rights which the patent owner considers its own through the claims appended to the patent. The publication of a description of the invention in conjunction with the claimed limits of the invention provides the public with notice of the patent owner's affirmative rights to the invention.

The process of invention generally starts in the transcribing of ideas that may, or may not, result in an advance or a solution to a recognized problem. Once an inventor is satisfied that the development has attained the desired level of usefulness, a summary of the inventive concept may be prepared. From this summary, including any appropriate data or laboratory work, an application for a U.S. patent may be written. The patent application is generally written by, or at least comes under the supervision of, a patent attorney or patent agent with one or more inventors. The participation of the inventors ensures that the patent application describes the invention in complete detail. The patent application should include the broadest definition of the invention and provide the best forms in which the invention can be practiced, applied, or used. The patent application should also have an explanation of the invention that describes it in terms considered definite by the inventor and that are used in the relevant technology. This terminology will enable others to understand and practice the disclosed invention.

Once the patent application is complete and the inventor has made a formal declaration of inventorship, the application is filed with the U.S. PTO. In the U.S. PTO, the application is the subject of a thorough, formal, and substantive examination by a patent examiner. Once the patent examiner is convinced that the patent application satisfies the statutory requirements provided for under the laws of the United States, the patent application will be allowed to issue as a patent. Issuance takes the form of a publication provided by the U.S. Government. The publication of patents occurs only on Tuesdays that are not federal holidays. At the time of issuance, the patent is assigned a number and made public in a form which allows all interested parties to obtain access to it.

The term of a patent depends on the date on which the application for patent is filed with the U.S. PTO. Patents filed and issued before June 8, 1995 had a term which is the longer of 20 years from the filing date or 17 years from issuance. The filing date is the earliest filing date relied on by the applicant. Patent applications filed before June 8, 1995 that issue after that date also have a term which is the longer of 20 years from filing or 17 years from issuance. Any original or follow-on patent application, ie, continuation, divisional, or continuation-in-part applications, filed after June 8, 1995 have a term of 20 years from filing, once it is issued as a U.S. patent.

The Nature of Invention

Invention may result from many different types of scientific or engineering efforts and advances. However, invention can also arise through the simple application of an idea that improves, refines, or otherwise modifies something that had been done previously. The simplest and most common area in which invention arises is in the development of products.

The nature of product development is such that it consists more of a process than a single discrete event. As a result, the objective, eg, developing a high cleaning detergent that is safe to the environment, may take place over a series of steps, rather than occur in one single, identifiable action.

The process of developing a product may result in one large breakthrough that could be considered a broad invention. This breakthrough may result in a new product that is useful and has many of the benefits that the inventor desired at the outset of the developmental project. However, the product still may or may not be suitable for commercial introduction or various other intended applications. As a result, further efforts may need to be expended toward refinement of the product so that it may take on its ultimate commercial form. Each of these potential refinements may also represent one or more patentable inventions that, while narrower in their intended usefulness than the original product, are still commercially valuable in their own right. In the search for improvements, refinements, and further solutions, invention thus may result either from a developed research effort or through the simple discovery of a solution to the original problem which is arrived at completely outside of the research context.

The resulting discoveries may provide a broad range of solutions or products. For example, invention may result from basic research and development efforts directed toward products which are essential to basic commercial efforts. Alternatively, invention may result in products or applications which add value to basic commercial products that are already in existence. Inventions may also be used to assist an individual or company in commercial efforts toward developing a defensive posture in any given marketplace. When patented, applications may also provide an extended opportunity to license or market the patent without the actual production of a product by the inventor.

READING A PATENT

Reviewing patent documents requires the skill of understanding the significance of what is being disclosed. Legal counsel should always assist in interpreting the legal effect of any patent on commercial activity. However, a patent attorney or agent often must seek the assistance of technical personnel to gain a full understanding of the technology disclosed and claimed in a given patent. Further, an understanding of the form, content, and function of the various sections of a U.S. patent assist the nonlawyer in understanding the commercial importance of any issued patent.

An abridged copy of U.S. Patent No. 5,131,727 is provided in Figure 1 to illustrate the elements of an issued U.S. patent. The cover or front page of a U.S. patent (Fig. 1a) must follow the form requirements placed on issued patents by the U.S. PTO. Specifically, the front cover discloses the inventor in two locations, A and C. The first named inventor is generally used as a head note, A, for the patent. A given patent may often be referred to in an informal sense by this inventor's name.

United States Patent [19]
Johnson (A)

[11] Patent Number: 5,131,727
[45] Date of Patent: Jul. 21, 1992

[54] AERODYNAMIC WHEEL COVER (B)
[76] Inventor: Harold M. Johnson, 2903 Legion Ave. North, Lake Elmo, Minn. 55042 (C)
[21] Appl. No.: 614,861 (D)
[22] Filed: Nov. 16, 1990 (E) (F)
[51] Int. Cl.⁵ B60B 7/02
[52] U.S. Cl. 301/37 P; 301/37 SA
[58] Field of Search 301/37 R, 37 P, 37 SA

[56] (G) References Cited
U.S. PATENT DOCUMENTS
3,565,489 2/1971 Eirenberg et al. 301/37 SA
3,847,443 11/1974 Laurion 301/37
4,209,230 6/1980 Perkins 301/37 SA X
4,620,749 11/1986 McEachern 301/37
4,660,893 4/1987 Huntzinger 301/37 SA
4,681,647 7/1987 Kondo et al. 156/175
4,682,821 7/1987 Strazis 301/37
4,712,838 12/1987 Berg et al. 301/37
4,729,605 3/1988 Imao et al. 301/104
4,732,428 3/1988 Monie 301/63
4,741,578 5/1988 Viellard 301/63
4,836,615 6/1989 Berg et al. 301/37
4,969,693 11/1990 Molson 301/37 R
4,978,174 12/1990 Nosler 301/37 R

FOREIGN PATENT DOCUMENTS
168747 1/1921 United Kingdom 301/37 SA
OTHER PUBLICATIONS
"Properties Cryovac D-940 Film," #S-06-20, Form 1584, W. R. Grace & Co.
"Cryovac D-940 Film," #S-06-20, Form 1583, W. R. Grace & Co.
"Mylar® LB, Summary of Properties" Form 195067A, DuPont.
"Mylar® HS, Summary of Properties," Form 194,150A, DuPont.
"Mylar® OL," Form H-36043, DuPont.
"Mylar® LBT-2, Summary of Properties," Form H-3336048, DuPont.
"Mylar® For Packaging, Summary of Properties," Type MMC, Form PBH, DuPont.

"Mylar® MB-P, Summary of Properties," Form 198116A, DuPont.
"Mylar® M-34, Summary of Properties," Form 202804A, DuPont.
"Mylar® For Packaging, Summary of Properties," Type 50 M-35, Form PBH, DuPont.
"Mylar® For Packaging, Summary of Properties," Type 50 M-44E, Form H-02951, DuPont.
"Mylar® M-45, Summary of Properties," Form 202805B, DuPont.
"Mylar® For Packaging, Summary of Properties," 75 M-45, Form PBH-6, DuPont.
"Mylar® For Packaging, Summary of Properties," Type 48 MM-20, Form E-71065, DuPont.
"Mylar® For Packaging, Technical Information Service, Type 48 MM-20," Form E-71066, DuPont.
"Mylar® For Packaging, Summary of Properties," Type 50 OL, Form PBH, DuPont.
"Mylar® 50 OL2, Summary of Properties," Form H-35232, DuPont.
"Mylar® For Packaging, Summary of Properties," Type 75 OL, Form H-02955, DuPont.
"Mylar® 75 OL 2, Summary of Properties," Type 75 OL 2, Form Jan. 30, 1991 PBH, DuPont.
"Mylar® 100 OL, Summary of Properties, Type 100
(List continued on next page.)

Primary Examiner—Russell D. Stormer (H)
Attorney, Agent, or Firm—Merchant, Gould, Smith, Edell, Welter & Schmidt

[57] ABSTRACT (I)
An aerodynamic wheel cover which includes a two sided circular cover, having an outer edge and an inner edge, and a central aperture. The wheel cover may be affixed by any number of adhesives deposited on one side of said aerodynamic wheel cover adjacent the wheel cover outer edge. The present invention also discloses a method of affixing the wheel cover of the present invention to wheels, the resulting wheels and vehicles.
29 Claims, 3 Drawing Sheets (J)

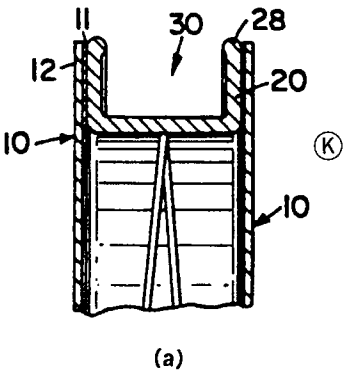


Fig. 1. The elements of an issued United States patent (a-c).

Once the patent is issued, the inventor is referred to as the patentee. The first named inventor, if there is more than one, is printed prominently in the upper left-hand corner of the front page of the patent, A. All of the inventors or patentees are listed beneath the invention title, B, along with the inventors' full names, addresses, and citizenship if other than the United States (3).

The title of the invention, B, is generally written so as to use the shortest possible accurate description of the invention described fully in the patent and found in the claims. The patent application number, D, and filing date, E, are printed beneath the title, B. The application number, D, and filing date, E, are important because the patent application filing date may be used to eliminate other publications of third parties that might be used to limit the legal scope of the applicant's rights.

Also printed on the front page of the patent is a coded classification listing, F. This coding is complex and largely unnecessary to a lay person's understanding of a patent. This classification stems from the specific technology area to which the patent application was assigned during processing in the U.S. PTO. The classification also results from the search or review of prior patents completed by the Patent Examiner.

Apart from the technical classification information, F, the front page of the patent also contains a listing of publications or references cited during examination, G, including "United States Patent Documents," "Foreign Patent Documents," and "Other Publications" such as trade literature, journal articles, and product descriptions.

The front cover of the patent generally also identifies the U.S. Patent Examiner who reviewed and allowed the patent application, as well as the patent attorney, agent, or firm who worked with the Patent Examiner on the application, H.

Also provided is an abstract, I, which describes the invention, specifically highlighting its most valuable properties and distinguishing features. By doing so, the abstract assists those searching for prior patents which disclose developments relevant to an invention or patent application presently under examination. Another aid to patent searchers is the listing of claims and drawing sheets, J. A representative drawing, K, may also often be found on the front page of the patent, if figures are provided by the inventor. Figures or drawings are not required to receive a patent. However, where figures are essential to a full and complete understanding of the invention, they must be included. Further, the figures should show those elements of the invention which are found in the claims.

Within the body of the issued patent, the title, B, now L, is generally repeated to maintain clarity (Fig. 1b). A field of invention, M, is then provided. The field of invention should direct the reader to the general area of technology to which the invention relates, and to specific improvements in the

identified areas of application. Generally, the field of invention is a fairly brief statement which allows the U.S. Patent Examiner to determine which technological area of the invention is the appropriate one.

1

Ⓐ AERODYNAMIC WHEEL COVER

Ⓜ FIELD OF THE INVENTION

This invention relates generally to light weight covers used to increase the aerodynamic properties of wheels. More specifically, the present invention relates to covers for wheels capable of creating an aerodynamic effect by reducing the friction or drag across the surface of a wheel created by air flow.

Ⓝ BACKGROUND OF THE INVENTION

Since the creation of wheels, man has sought various mechanisms to assist wheels in turning at a higher rate. Hubcaps or wheel covers have been used for years to provide ornamental decoration for wheels. Generally, mechanical means of attachment such as friction clips and screws have been used to attach the hubcap to the wheel rim. Traditionally, these approaches have not been used on two wheel vehicles such as motorcycles and bicycles as there is no effective means for attaching hubcaps to an axle that extends beyond the planar, cross-sectional thickness of the wheel rim. Moreover, the weight of hubcaps or wheel covers generally used in the automotive industry are not suitable for motor or human powered cycles.

Recently, composite wheels have become popular. While composite wheels may eliminate the use of spokes, they are costly and do not necessarily reduce the weight of the wheel or the energy necessary to initiate revolution. As a result, composite wheels do not always provide an adequate alternative to wheel covers as they may not be readily applicable to all uses in which spoke wheels may be found. Traditional spoke wheels are still the predominant wheel system for most two-wheeled vehicles.

In the past, various systems have been proposed for covering spoke wheels. For example, U.S. Pat. Nos. 4,712,838 and 4,836,615 to Berg et al discloses a clip-fastened disc cover for spoke wheels, which generally consists of a fabric or plastic cover having a hoop of semi-rigid material in a peripheral pocket of the cover. McEachern, U.S. Pat. No. 4,620,749, discloses a fabric or polymeric wheel cover which generally consists of a porous cover, having a central aperture for the wheel hub held on wheels by tension engagement with an opposing cover.

Laurion, U.S. Pat. No. 3,847,443, discloses an ornamental wheel element which is designed to fit between the spokes, inside a wheel. Strazis, U.S. Pat. No. 4,682,821, discloses a semi-rigid, tension attached bicycle wheel cover assembly intended to improve the aerodynamic efficiency of bicycle wheels. Monte, U.S. Pat. No. 4,732,478, discloses a streamlined wheel for bicycles which comprises two hollow half shells which are joined to create a rim for support of a tire. Imao et al, U.S. Pat. No. 4,729,605, and Viellard, U.S. Pat. No. 4,741,578, discloses spokes and wheel components useful in composite wheels.

However, these systems fail to disclose an inexpensive means of easily improving the aerodynamic properties of a spoke wheel with minimal manual effort. As can be seen, while any number of alternative wheel covers are available, these systems have certain shortcomings which have not yet been satisfied by the art.

2

Ⓞ SUMMARY OF THE INVENTION

The present invention provides a wheel cover which improves the aerodynamic properties of the hub-rim-spoke wheels. The wheel cover of the present invention is preferably made of a light weight synthetic or natural polymer, fabric or paper film which is adhesively applied to a wheel through simple manual application. In turn, the aerodynamic wheel cover of the present invention may also be easily removed and replaced to allow on-road repairs of spoke, hub, or rim.

Preferably, the wheel cover of the present invention may be easily applied without extended mechanical or manual effort merely by adhesively applying the cover to the wheel spokes or rim. Once in place, the wheel cover may be shrunk to size so as to provide a tightly fit cover.

In accordance with the present invention there is provided, an aerodynamic wheel cover comprising a two sided circular cover having an outer edge and an inner edge. The inner edge of the wheel cover borders a central aperture. Adhesive means is deposited on one side of the aerodynamic wheel cover adjacent to the wheel cover outer edge. Also disclosed are methods for applying the vehicle wheel of the present invention, and the resulting wheels as well as wheeled vehicles.

BRIEF DESCRIPTION OF THE DRAWINGS Ⓟ

FIG. 1 is a side perspective view showing the wheel cover of the present invention in application on a bicycle wheel.

FIG. 2 is a side plan view of one embodiment of the wheel cover of the present invention shown in FIG. 1 with the wheel cover applied to a spoke wheel rim.

FIG. 3 is an alternative embodiment of the wheel cover of the present invention shown in FIG. 1 with the wheel cover applied in this instance to facilitate friction fitting the wheel cover between the interior of the rim and a later applied wheel tire (not shown).

FIG. 4 is an alternative embodiment of the wheel cover of the present invention shown generally attached to a spoke wheel at the spokes.

FIG. 5 is a cut away view of the wheel covers of the present invention shown in FIG. 4 taken along lines 5-5.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT Ⓠ

The present invention discloses a wheel cover, a method of applying the wheel cover, and resulting wheels and vehicles.

The Wheel Cover

Turning to the figures, wherein like parts are designated by like numerals throughout several views, there is shown in FIG. 1 an aerodynamic wheel cover 10 in one environment of application, that is, applied to the wheels of a bicycle 1. The aerodynamic wheel cover 10 generally comprises a two sided circular element having an outer edge 12 and an inner edge 14. The inner edge 14 borders a central aperture 16. Deposited at the outer edge of one side of the aerodynamic wheel cover is an adhesive 11. FIG. 2. The adhesive 11 may generally be positioned adjacent to the wheel cover outer edge 12 to assist in affixing the wheel cover 10 to various elements of the wheel.

In accordance with the present invention, the wheel cover disclosed in FIGS. 1-5, generally functions to

(b)

Fig. 1b. (Continued)

A description or explanation of the background of the invention, N, may also be provided by the inventor. This background section discusses previous developments of inventors working in the same area of technology and may also list publications or patents that have discussed these developments and predate the filing date of the patent application. The background section may also point to deficiencies in the prior developments that the inventor intends to overcome.

To complement the discussion of problems and prior publication in the background of the invention, N, the inventor may generally provide a summary of the invention disclosed in the instant patent. The summary of the invention, O, should provide an explanation of the invention in the broadest and simplest terms and should also discuss how the invention disclosed in the patent solves problems remaining in prior work in this area of technology.

The patent should also provide a brief description of any drawings or figures, P. This brief description is often given in the technical terms used by engineering draftsmen to explain the various views illustrated in the figures.

The next section of the patent is titled "The Detailed Description of the Preferred Embodiment", Q (Fig. 1b), often a multipage work serving several functions. First, the detailed description should provide an illustration of the invention in both its broadest or simplest sense and in its most preferred sense. Any elements of the invention that the inventor believes are crucial to the success or performance of the invention must also be included within this description. Further, this description should provide an explanation of the invention that is definite and illustrative, so as to allow persons having nothing but the patent before them to practice or use the invention in the manner intended. This description should be understood by those who work in the area that covers the subject matter of the patent.

Elements of Q often include a detailed explanation of the various elements of the invention comprising the function of those elements, a written description of those elements, and an analysis of the elements that relies on any figures present in the patent application. The Detailed Description of the Preferred Embodiment, Q, may also include one or more working examples, R, especially if the invention is related to chemical technology (Fig. 1c). That is, in cases relating to chemistry, biochemistry, and chemical engineering, working examples are more often included than not. These working examples may serve any number of functions, including illustrating the formulation, applicability, and performance of the invention. Working examples may also be

used to illustrate how the invention is distinguishable from those inventions previously developed and patented. As such, these working examples may include data such as adhesion and cohesion performance for adhesives, disinfecting and sanitizing efficacy for cleaners, or data on chemical and physical properties for polymer systems.

11

examples demonstrated that a cover can be constructed that will be permeable to air. This is an important aspect to consider for bicycles where cross winds can impart a substantial lateral force that can create handling problems for the cyclist.

(R) WORKING EXAMPLE 4

A triangular shaped wheel cover with the center cut out to provide hub access was then applied to a wheel. The cover was constructed from a heat shrinkable polyolefin film. Tape was attached to the apex points of the triangle. The tape liner was removed and the three adhesive sites were fastened to the spokes. As an identical complementary cover was then applied to the opposite face of the wheel in a mirror image fashion. The adhesive contact points were positioned to encapsulate the spoke on either side within the adhesive contact point. Heat was then used to shrink the covers and achieve a wrinkle-free condition. This example demonstrates that design can play a part in providing a stylish wheel cover that is capable of individualizing the bicycle to meet a wide variety of consumer tastes.

The foregoing specification, examples and data provide a basis for the understanding of the invention. The invention can be made on a variety of embodiments without departing from the spirit and scope of the invention. Accordingly, the invention resides in the claims hereinafter appended.

(S) I claim as my invention:

- 1. A heat shrinkable aerodynamic wheel cover comprising
 - (a) a two sided circular cover, said cover comprising a polyolefin material, said cover comprising an outer edge and an inner edge, said inner edge bordering a central aperture;
 - (b) adhesive means deposited on one side of said aerodynamic wheel cover, said adhesive means positioned adjacent to the wheel cover outer edge wherein said polyolefin material does not interfere with the mechanical operation of the wheel and has a tensile strength of about 200 to 25,000 psi.
- 2. The aerodynamic wheel cover of claim 1, wherein said cover thickness ranges from about 0.5 mils to about 125 mils.
- 3. The aerodynamic wheel cover of claim 1 wherein said adhesive means comprises an adhesive selected from the group consisting of velcro, adhesive tape, or an adhesive resin.
- 4. The aerodynamic wheel cover of claim 3, wherein said adhesive means comprises an adhesive selected from a group consisting of natural or synthetic thermoplastics, and thermosets.
- 5. The aerodynamic wheel cover of claim 4, wherein said thermoplastic adhesive comprise a pressure sensitive adhesive.
- 6. The wheel cover of claim 4, wherein said thermoplastic adhesives are selected from the group consisting of polyamides, polycarbonates, polyesters, polyolefins, polyvinyl acetates and combinations thereof.
- 7. The aerodynamic wheel cover of claim 4, wherein said thermoset adhesives are selected from a group consisting of epoxies, phenolics, isocyanates, cyanoacrylates, acrylics or combinations thereof.
- 8. The wheel cover of claim 1 wherein said cover has a thickness of about 1 mil to 60 mils.
- 9. The cover of claim 1 wherein said cover has a thickness ranging from about 3 mils to 15 mils.

12

- 10. The cover of claim 1 wherein said polyolefin material has a tensile strength ranging from about 300 psi to 15,000 psi.
- 11. The cover of claim 1 wherein said polyolefin material comprises polypropylene.
- 12. The cover of claim 1 wherein said polyolefin material comprises polyethylene.
- 13. A vehicle wheel comprising
 - (a) a wheel rim;
 - (b) a hub positioned within the circumference of said rim;
 - (c) a plurality of spokes extending inwardly from said rim to said hub; and
 - (d) at least one heat shrinkable aerodynamic wheel cover affixed to the wheel, said wheel cover comprising a two sided circular cover, said cover comprising a polyolefin material having a tensile strength of about 200 to 25,000 psi, said cover comprising an outer edge and an inner edge, said inner edge bordering a central aperture adjacent said wheel hub, and adhesive means deposited on one side of said aerodynamic wheel cover, said adhesive means positioned adjacent to the wheel cover outer edge wherein said wheel cover does not interfere with the mechanical operation of the wheel.
- 14. The vehicle wheel of claim 13, wherein said wheel cover outer edge is attached to the rim through said adhesive means.
- 15. The vehicle wheel of claim 13, wherein said wheel cover outer edge is attached by fixing said adhesive means to said spokes.
- 16. The vehicle wheel of claim 13, comprising a second heat shrinkable wheel cover positioned over the second side of the wheel.
- 17. The vehicle wheel of claim 16, wherein said first and second wheel cover outer edge is attached by fixing said adhesive means to said spokes.
- 18. The wheel cover of claim 13 wherein said cover has a thickness of about 1 mil to 60 mils.
- 19. The cover of claim 13 wherein said cover has a thickness ranging from about 3 mils to 15 mils.
- 20. The cover of claim 13 wherein said polyolefin material has a tensile strength ranging from about 300 psi to 15,000 psi.
- 21. The cover of claim 13 wherein said polyolefin material comprises polypropylene.
- 22. The cover of claim 13 wherein said polyolefin material comprises polyethylene.
- 23. A method of applying a heat shrinkable aerodynamic wheel cover to a vehicle wheel, said vehicle wheel comprising a rim and hub, said aerodynamic wheel cover comprising a two sided circular cover, said cover comprising a polyolefin material having a tensile strength of about 200 to 25,000 psi, said cover comprising an outer edge and an inner edge, said inner edge bordering a central aperture, and adhesive means deposited on one side of said aerodynamic wheel cover, said adhesive means positioned adjacent to the wheel cover outer edge wherein said wheel cover does not interfere with the mechanical operation of the wheel, said method comprising the steps of:
 - (a) stretch applying the aerodynamic wheel cover to a hub and rim wheel; and
 - (b) heat shrinking said wheel cover after application.
- 24. The method of claim 23, additionally comprising the step of applying a second cover to said wheel.

(c)

Fig. 1c. (Continued)

The final section of an issued patent is the claims, S. A United States patent is required by law to have at least one claim. The claims represent the legal definition and boundaries of the rights resulting from the patent grant. Patent claims are analogous to the legal description which one might find on a title to real estate.

When evaluating an issued patent for purposes of determining the patentability of a new invention, the entire patent must be considered. As a result, the figures, K, and The Detailed Description of the Preferred Embodiment, Q, are every bit as important to an issued patent as the claims, S. At certain times any one of these elements may become more relevant than another. For example, claims tend to be more relevant to determinations of patent infringement or violation. However, in determinations concerning the patentability of new inventions, the figures, K, and The Detailed Description of the Preferred Embodiment, Q, may be the most relevant aspects of any previous patent.

THE TECHNICAL SUBJECT MATTER OF PATENTS

A fundamental requirement for obtaining a patent is defining an advance, development, or invention which is within those classes of "subject matter" which the law of the United States regards as patentable. Two classes of patentable subject matter, ie, computer software and biotechnology, are the subject of relatively new and evolving law. However, other types of subject matter rest on fairly certain ground as to patentability. Examples of patents directed to various types of subject matter are described in the following.

Composition of Matter. This is the subject matter category into which many chemical and biochemical (and biotechnology) inventions fall. Composition of matter includes a compound, a mixture of compounds, or a reaction product stemming from a mixture of compounds. Inventions such as pharmaceutical products, herbicides, cleaning agents, adhesives, food products, and personal care products such as facial cleansers and shampoos are all commonly regarded as compositions. An example of a patent disclosing compositions of matter is U.S. Patent No. 5,157,128 titled "Certain Optically Active Substituted α,α -Dialkyl-Pyrrolidine-3-Methamines Useful as Intermediates" (1).

Article of Manufacture. An article of manufacture is an invention such as a two-headed tooth brush, an intravenous fluid bag, or an optical fiber "made" by a machine. One example of an invention which could be considered an article of manufacture is **U.S. Patent No. 5,241,990** titled "Irrigation Aspiration Valve and Probe for Laparoscopy" (2).

Machines. A machine is a device which is capable of manufacturing a product or completing a task such as removing hydrocarbon contaminants from silica and dirt. Examples of machines include an extrusion apparatus, a book binder, and a tractor. The **U.S. Patent No. 5,020,462** titled "Thermal Remediation Apparatus and Method" (3) discloses both a machine and a process.

Processes. Methods or processes represent patentable subject matter regardless of whether the invention represents a method of using or a method of manufacturing an article, composition, or device. Examples of methods are a series of steps for manufacturing urethane, the sequence for formulating a stable injectable pharmaceutical composition, the manner in which an electrical circuit board is assembled, or a method of treating a disease using a compound or composition. "Method of Making Metal—Film Laminate Resistant to Delamination," **U.S. Patent No. 5,112,462** (4), and "Clarification Process for Mining Liquors," **U.S. Patent No. 4,997,573** (5), are patents which disclose various processes.

Design. Ornamental designs are also a legally recognized class of patentable subject matter. The design must be embodied in an article of manufacture, such as a concrete masonry block or a sun screen for a car window. An example of a design patent is **U.S. Patent Design No. 334,420** titled "Chemical Detergent Block" (6).

Plants. Asexually reproducing plants, ie, those not propagated by means of seed, also represent a legally recognized class of patentable subject matter under U.S. patent laws. Additionally, the inventor must have discovered and asexually reproduced the plant that is to be the subject of the patent application. Plant patents are assigned a different series of numbers than the majority of patents discussed in the foregoing, such as **U.S. Plant Patent No. 3,360** titled "Peach Tree" (7).

Assigning the Subject Matter Class. Factors to consider when reviewing the "nature" of the invention may be summarized by the following questions:

- What is the technology area of the invention?
- How was the invention made?
- How is the invention used?
- Does the invention "do something" or is it "something that is done or made"?
- Does the invention have an evident usefulness, and if so, what is the ultimate usefulness of the invention?
- Regardless of whether the invention is useful, is it ornamental?

For example, if the invention has cleaning properties it may be a composition of matter, process, or machine, all of which are patentable subject matter. If the invention performs work, it may be a process, article of manufacture, or machine, which are all patentable subject matter. It may also be that the development of a new composition results in a composition of matter and a process of using the composition, both of which are distinct, yet patentable inventions. The various types of patentable subject matter are not mutually exclusive and may be disclosed in a single patent.

THE ORIGIN OF INVENTION

Invention results from the application of an idea or concept. The idea itself is generally not patentable. An application of the idea may be patentable if it falls within one of the categories of subject matter previously discussed. For example, the idea may be to increase friction or traction in road surfaces during any of a variety of weather conditions. There may be any number of ways to apply this idea. One example of an application would be the creation of grooves in road surfaces. Such channeling of a surface may be found to expedite the drainage of rainwater from the road surfaces. The concept of providing improved traction on road surfaces is certainly not patentable in and of itself. However, once applied, providing drainage channels on the roadway is certainly inventive and possibly patentable, depending on the previous solutions to the presented problem.

A further application of the concept may be found in the patterning of automobile tires to channel residue incident to contact with the road surface away from the automobile tire. Any of these applications of the central idea of providing a solution to increase road traction also may be patentable.

Although it is not always necessary, a practical application of a concept may move through a series of steps or stages. Indeed, the recognized pathway to invention involves at least two factors, ie, conception of the invention and reducing the invention to practice. Returning to the earlier example, the inventor may conceive an application such as modifying the chemical and physical structure of the automobile tires to improve traction. However, without further research the resulting tire may not completely solve the problem and may even create additional problems. For example, softer tires provide greater traction but may also wear more quickly. Investigations may be undertaken on the various levels of tire softness and rigidity so as to accommodate the varying types of weather in which the tire is to be applied. Research may be done to determine the applicability of the various types of synthetic and natural rubbers available for use. Research may also be undertaken to alter polymerization processes so as to produce tires having varying physical properties or design patterns on the face of the automobile tire that may have varying effects on road residues. All of these efforts are directed toward reducing the invention to practice.

The initial research effort may prove to be a broad spectrum of applications or solutions to the original problem that in turn provide any number of inventions. When efforts move toward reducing the invention to practice and refining the invention so that it proves to be commercially marketable, certain applications may prove to be unfeasible or commercially impractical. As a result, only one application, eg, the creation of a given pattern on the surface of the automobile tire, may ultimately prove commercially marketable. However, all the solutions which are developed and considered over the research and development process may comprise inventions that are worthy of disclosure and claiming in a patent. An application which is not commercially viable today may become viable within the seventeen-year lifetime of a patent.

Unlike the common practice occurring in other countries, in which award of patent rights is based on the date on which a patent application is filed, in the United States the patent grant is based on the first date of invention. To be an inventor in the United States, an individual must contribute to conception of the invention, and may contribute to reduction of the invention to practice. Although the creation of an advance, development, or application may be conceived by one given individual, it often is the case that the act of invention is the work of many individuals, especially in a commercial context. Accordingly, inventorship questions often arise.

Inventorship. Those who may deserve to be considered inventors include all those who have contributed to conception of the invention. Further, those who have provided contributions which would be considered something above and beyond textbook knowledge in the reduction of the invention to practice may also deserve to be listed as inventors.

The legal guidelines which direct inventorship determinations are some of the most stringent and complex in modern patent law. Many factual and legal analyses may justify the listing of an individual as an inventor, and certain levels of contribution do merit this designation. However, under U.S. law an individual is most probably *not* an inventor if: (1) that person has merely supervised laboratory operations without providing any general or specific contribution to the project at hand; (2) that person has functioned in a capacity that is substantially, if not wholly, directed by a supervising scientist or engineer; or (3) that person has used a level of skill that would be expected in the ordinary routine of the production, evaluation, or analysis of that which later embodies the invention.

Developing the Record of Invention

Developing the record of invention is an important, if not a fundamental, point in the process of securing protection over the invention. Generally, there are two stages in the development of the record of invention. The first stage is the laboratory or experimental work that is done in conceiving the invention and applying it within the intended area of use. This may take place over a period of weeks, if not months, with continual or intermittent work toward the ultimate production of an advance, development, or application which solves various problems. The second stage of developing a record of invention is the actual process of defining the invention along with the noting of any events or facts that may limit the invention.

The Laboratory Notebook Page. Most engineers, scientists, and technicians make a record of their work. A common form of record

keeping is the use of a laboratory notebook (Fig. 2). Experimentation is usually undertaken with an intended objective. Recording this objective often assists in illustrating the purpose or quality of an invention. The laboratory notebook should reflect elements, parameters, conditions, and thoughts that were material factors in the completion of a given experiment. For example, in the formulation of a chemical mixture the constituents of the mixture should be defined by their accepted chemical name (and trade name if available) and concentration. Further, other parameters relevant to the successful formulation of the mixture such as mixing time, temperature, volume, and pressure should also be detailed. Parameters that were thought to be completely irrelevant may become relevant once the scientist has reflected upon notations made over the course of repeated experiments. Testing of quality or efficacy should also be recorded in detail. Once the experiment is completed, the scientist, engineer, or technician who has undertaken the work should confirm completion of the work by signing and dating the record. The record should also be witnessed by at least one other person who reads and understands it, and did not take part in the experimentation.

Laboratory Notebook No. 1412

Page No. 37

Iwana Patent Corporation

Project: Diet Syrup

Date: 3/11/94

Objective: To formulate a storage-stable, phase-stable syrup for baked breakfast goods which is sugar-free.

Constituent	wt %
Maple syrup	10
Sugar syrup	60
High fructose corn syrup	20
Flavoring	0.1
Cellulose gum	5.0
Water	Balance

The ingredients were mixed at 100 degrees C for 15 minutes until a homogenous formulation was created. The product was then frozen at 0 degrees C for 5 h and then thawed. This cycle was repeated 10 times without any evident precipitation separation.

Completed by: Shiela Oswald 3/11/94

Read and understood by: Steven Jay 3/12/94

Fig. 2. Laboratory notebook page.

Elements that should be considered in developing a laboratory or experimental record are as follows:

- Is the record found in a notebook which has been bound?
- Are the record entries in chronological order, with an emphasis on avoiding the skipping of pages?
- Is there a full record of a chronology of each experiment, including the starting date of every experiment plus each day's entry?
- Has the objective of each experiment been stated within the record?
- Have all essential facts been recorded so that if abbreviations are used, they are unambiguous?
- When more than one page is required, are references provided to previous and subsequent pages?
- If a standard or routine procedure is being used in the experimentation, has that procedure been referred to by location or full description?
- Is the record void of any comments concerning patentability?
- Is the record complete so as to provide conclusions and evaluations of the results stemming from the experimentation?
- Is the record complete in providing analytical data or referring to the place where analytical data can be found?
- Is the record unambiguous to the extent that unused portions of record pages are lined through and that there are no erasures or backdated entries?
- Has the recording individual used permanent ink?
- Has the record been witnessed promptly by at least one individual who did not participate in the experimentation, one who can read and understand the description, and one who preferably, but not necessarily, observed the experiment?

All of these factors should be considered when evaluating the quality of laboratory notebook entries. These entries may otherwise never be considered until they are the subject of a legal contest, at which time quality review may be too late.

The Record of Invention. The second phase of developing a record of the invention is to condense the record into a summary form which serves several purposes. Specifically, the record of invention establishes a date of invention through attached copies of notebook records, spectra, and the like which all prove that the invention has in fact been conceived and reduced to practice in some form having practical utility.

Along with other elements of the invention, it is good practice to include within the record of invention any first written descriptions or drawings of

the invention. It is also prudent to attach photocopied notebook pages that evidence this development. Another helpful part of the record of the invention is to list the names of individuals who worked on the invention and enumerate their respective contributions.

One further component of the record of invention is a list of any uncovered publications or patents which are relevant to the invention. Such a listing should also include any disclosures made by any of those who worked on the invention to other parties inside or outside the organization. All inventors should sign the record of invention. At least two witnesses who are not inventors should also read and understand the record of invention so that they can sign and date this document.

The importance of an accurate and complete record of invention cannot be underestimated. The record of invention should serve as the basic document for establishing the date of conception and reduction to practice of the invention. The U.S. PTO issues patents to those who are first to invent. In a contest over inventorship, any available record of invention is submitted to the U.S. PTO to establish proof of an inventor's rights. As of January 1, 1996, any inventor from a country belonging to the World Trade Organization may use such evidence before the U.S. PTO. Previously, this type of proof could be relied upon only if the activity, documented in the notebook, record, etc, was undertaken in the United States. Similarly, activity undertaken after December 8, 1993 in Mexico or Canada may also be relied upon to prove inventorship.

The following provides a meaningful but not all-inclusive checklist of factors to consider when completing a record of invention.

- Has the technology and commercial field to which this invention relates been identified?
- Has a search of the prior patents and literature been done and copies of the search results or the publications been attached?
- Have the most relevant prior patents and literature been determined?
- Is the invention an improvement over prior patents and literature and has the nature of the improvement that the invention presents been identified?
- Has explanation in detail of how the prior patents and literature have been improved and what specific problems were solved been provided?
- Has the unexpected or surprising property provided by the invention been identified?
- Have the essential elements of the invention been identified?
- For each essential component of the invention, have the following been identified: its function? a general definition of the component? a specific list of materials which may be used for each component? what is the preferred material? limits, boundaries, ranges at which the material is useful? preferred limits, boundaries, ranges?
- Have noncritical components of the composition been identified?
- Has the intended commercial embodiment of the composition been identified?
- Have reasonable alternatives for each component been identified?
- Has the invention been described in enough detail to enable someone with skill in the art to make it?
- Are copies of the relevant lab notebook pages attached?
- Were any graphs or design experiments completed? If so, are they attached to the disclosure?
- Are the data sheets for each raw material attached to the disclosure?
- Have any tests been done to support the claim that the invention is better than that which is disclosed in the prior patents and literature? Are they attached and identified as such?
- What is the nature of these tests? Are the protocols for the tests attached or identified?

Determining the Scope of the Invention

Once the record of invention has been written, an evaluation of the invention should be undertaken. A careful evaluation of the record of invention is usually best completed by a committee of individuals from technical, commercial, and legal disciplines. It is important to include the viewpoint of those scientists working in the field, those commercial or sales people who will be responsible for selling any products which stem from the invention, and those individuals who may be able to offer a legal opinion given the insights of commercial and scientific personnel.

First the committee should consider the technical merit of the invention. Specifically, is it reasonable from a scientific or engineering standpoint? Further, is there a clear advance in technology that has not been previously undertaken or achieved by another party?

It is also important to ascertain the commercial significance of the invention. Although the invention may provide a measurably large advance in technology, science, or industry, it may not provide an easily producible commercial vehicle or product. Alternatively, the invention may be easily produced as a commercial product, but that product may have limited relevance to the overall commercial strategy or plans of the organization.

The legally trained member of the interdisciplinary committee should provide insight as to the significance of the technological advance and as to whether any commercial product ultimately derived from the invention could be protected by an issued patent. Another important function of this person is to determine the scope of the invention based on preceding events, publications, or activities which may have otherwise limited the breadth of the invention. To this end, U.S. law requires that an invention satisfy a number of prerequisites or requirements before issuing a patent: novelty, nonobviousness, utility, and disclosure.

NOVELTY

A fundamental statutory prerequisite to patentability is novelty. A lack of novelty occurs when each and every element of the invention, as it is claimed, is found in a single disclosure which occurs before the date of invention. Such a disclosure may occur in any of a number of forms. To be an adequate disclosure, it should be catalogued or inventoried as a book might be in a reference library and open to public dissemination. The novelty requirement presents the inventor with an extensive list of "cans" and "cannots." Unfortunately, the natural course of research and development often leads to activities which are much more readily categorized as "cannots" than "cans." Ultimately these activities may even proscribe the issuance of a patent if an application is not filed in a timely fashion.

- Questions that should be considered when determining whether an invention is novel include the following.
 - Was the Invention Known or Used by Others?** The invention cannot have been publicly known or publicly used by others or the subject of a patent or publication anywhere in the world prior to the applicant's actual invention date. If someone other than the inventor has published a journal article, received a patent, or used the invention publicly, the inventor will not be able to receive a patent on the invention.
 - Was the Invention Used, Sold, or Advertised For Sale?** The invention cannot have been the subject of an offer for sale, public use, or a patent or publication published more than one year prior to the filing date of the inventor's patent application by the inventor or any other party. This rule means that an inventor may lose the right to patent an invention even though pursuing a normal and expected course of events toward placing the invention in the commercial market. From that point in time in which the inventor discloses the invention to the public, either by advertising the product, publishing an article on the product, placing it on sale, or by allowing a public use of the invention, the inventor has one year to file an application. Otherwise, any right to a patent stemming from the invention will be lost to the public domain.
 - Was the Invention Abandoned?** The invention cannot have been abandoned. An invention may be abandoned either expressly or impliedly. For example, abandonment may occur when an inventor expressly disclaims the invention by dedicating it to the public. Abandonment may also occur if a patent applicant fails to complete the examination of a patent application pending within the U.S. PTO during the time periods set for completion. The publication of an article disclosing the invention may be an abandonment where the inventor does not file an application for a patent within one year.
 - Has the Invention Been the Subject of a Prior Foreign Patent?** Although one cannot be sure when, or even whether, a patent will issue from any application, it is good practice to make sure your U.S. patent application is filed within 12 months of the first foreign filed application. The patenting of an invention in another country by the inventor or another party will preclude the issuance of a patent on the same invention in the United States.
 - Has the Invention Been the Subject of a Prior U.S. Patent?** A previously filed U.S. patent application or issued patent disclosing the

same invention and originating with another inventor may be sufficient to deprive the second inventor of the desired novelty. An inventor who is the first to file a patent application in the U.S. PTO will retain priority of invention and is entitled to a patent over another patent applicant who subsequently files for a patent. Events which may destroy novelty are often also referred to as "prior art," given their nature as an earlier event which is relevant to the technical art.

Other examples of prior events or prior art which may destroy novelty are as follows:

- Graduate school dissertations such as Masters or Ph.D. dissertations.
- Abstracts of meetings of technical organizations.
- Approved or published grant proposals, as well as status reports on ongoing grants.
- Published articles in the popular press.
- Prior products manufactured by the client.
- Prior company literature related to the invention.
- Prior publications of the inventor in the area.
- Research in the technical area of the invention.
- Companies researching in the area of the invention.
- Any patents in the area.
- Presentations made by the inventor or others in the area of the inventions at trade shows, conferences, etc.
- Post-sessions disclosing the invention or other materials related to the invention.
- Demonstrations of the invention to customers or other third parties.
- Any patent applications filed by others in the area of the invention before and after they become publicly available.
- The publications and work of any institutes, associations, or industry groups.
- The trade names and information on competitive products in the commercial area of the invention.
- The contents of all prior filed applications by the inventor with others or by others in his or her organization which are related to the invention.

Economic reality dictates that the invention must eventually be commercially exploited. Experimental trials are a natural follow-on to laboratory work and are often necessary to further refine or otherwise reduce the invention to practice. Although the trial may not preclude the subsequent filing of a patent application on the invention, such experimental trials should be reviewed in advance to determine the effect they may have on filing for patent protection. For example, in some circumstances trials made for purposes of gaining further experimental data on the invention may be perfectly acceptable. In addition, even if a trial is made in public, it may be the case that this trial does not extinguish the novelty of the invention. The application of further refinements to the invention, the facts surrounding the trial, and the ultimate timing of the filing of the patent application may all be determinative of whether or not the novelty of the invention survives.

In any event, the highest importance should be accorded to the coordinating of events which may affect the novelty of the invention. Careful consideration should be given to the importance and timing of promotional events. It is often the case that patent applications can be filed and drafted well before announcements occurring at technical conferences. Further, technical publications often have an extended lead time before they are actually published. In any instance, the filing date of a patent application retains extremely great importance, being a determining factor in the timing of any disclosure.

NONOBVIOUSNESS

The grant of a patent is also dependent on whether the advance, application, development, or invention is obvious. If an invention is obvious, it is not patentable. The legal qualification of obviousness is a very difficult concept to understand. Although all the elements of an invention may actually be published, if they do not appear together in a single publication, then the invention is generally still novel. However, if the publications may be read in combination to disclose all elements of the invention, the invention may be considered obvious and not patentable.

An initial determination on the degree to which an invention may be "obvious" can be obtained by answering the following questions:

- What do prior patents, publications, and public activity disclose relative to the invention?
- What are the differences between all of this prior activity and the new invention?
- Would the skilled technician, engineer, or scientist consider the newer invention unexpected or surprising in view of this previous work?

However, even if there is some disclosure of the invention in the prior activity, the law of patents in the United States requires a high level of detail concerning the invention. A summary of factors to consider in establishing that an invention is not obvious is as follows:

- The results achieved by the invention are new, unexpected, or superior.
- Up to now, the techniques used in the invention were unworkable.
- Up to now, problems solved by the invention were not solvable.
- The invention has attained commercial success.
- The problem solved by the invention was never recognized before.
- An element of a prior invention has been omitted without loss of capability.
- Prior teachings lack any suggestion that the reference should be modified in a manner required to meet the claims.
- Up to now, those skilled in the art never appreciated the advantage of the invention, although it is inherent.
- The prior patents and or literature are inoperative.
- The prior patents and or literature are vague, conflicting, or very old and therefore are weak and should be construed narrowly.
- The invention has been licensed.
- The invention has been given an award or recognized in a professional publication.
- The invention has been copied by an infringer.
- The result achieved by the invention is greater than the results achieved by any of the individual prior teachings.
- Scientists, engineers, and technicians would find it physically impossible to combine the prior teachings to produce the invention.
- If combined, the prior teachings would produce an inoperative combination.
- The prior teachings themselves teach away from the invention.

UTILITY

Aside from designs and plants, inventions are required to exhibit usefulness or utility to be patentable. In fact, issued patents for processes, machines, compositions, and articles are often commonly referred to as "utility" patents. Depending on the nature of the technology, a single assertion of utility may suffice. In other cases, such as in the field of biotechnology, a more elaborate demonstration of utility may be necessary. Although utility may be supported by an assertion of use, application, or benefit, the assertion must be accurate and credible to ensure the enforceability of any patent relied upon to cover the invention.

An inventor may establish utility by providing several working examples which disclose preparation, application, and even some or all of the benefits of the invention. Utility may also be substantiated by merely disclosing several applications for the invention. One method of determining the breadth or scope of an invention is to define the invention by only those elements essential to performing the intended task. This definition should then become the broadest claim of the patent application.

DISCLOSURE

An additional statutory requirement is that of disclosure. A patent must provide the public with a disclosure which is enabling, definite, and shows the best mode for practicing the claimed invention.

Enablement. The patent has to enable any person reading the disclosure who has skill in the relevant technical area to make and use the invention. The enablement requirement mandates that the applicant provide a description of the process of manufacturing given invention. Also, the patent provides an adequate description of the process of using the invention. This enables a person of adequate skill in the technical area to which the invention pertains to be able to make and use the invention without undue experimentation. In applying the disclosure to make or use the invention, a certain level of experimentation is allowed. However, the experimentation cannot be undue, requiring the reader essentially to recreate the invention through extended and potentially unsuccessful guesswork.

The disclosure requirement provides that the patent be a teaching document, and enhance the breadth of knowledge held by the public. By increasing the breadth of knowledge within the public sector, a given patent facilitates further technological development and growth, which in turn results in the issuance of additional patents.

Problems with enablement arise when the patent fails to provide an adequate disclosure of parameters or materials for use in producing or performing the invention. The enablement requirement may, however, be satisfied by relying on and referencing a particular level of experience or knowledge in the given field of technology and incorporating that reference directly into the patent application.

Definiteness. Adequate description or definiteness requires that the patent claims provide an outline of those elements which are integral to the application's invention. In turn, the specification acts as a dictionary wherein the reader can interpret and understand the elements in the patent claims. Complementary to the requirement of definiteness is the requirement that the application must disclose the entire invention. The applicant cannot make a claim of right to the invention where essential elements of the invention are not disclosed in the patent.

The definiteness requirement serves notice to potential infringers as to the exact boundaries of the patentee owner's rights. Thus, a patent provides a record of what the inventor has brought to the technological field, and also provides other parties with notice as to what conduct is permissible in view of the patent claims.

Best Mode. The patent applicant must disclose the best mode of practicing the invention known to the inventor at the time the application is filed. Concerns over best mode often arise when a patent applicant seeks patent protection for an invention but, at the same time, desires to keep as a trade secret one aspect of the invention necessary to the production of a commercial product. This action denies the public access to this information and undermines the policies of the patent system. This would effectively deprive the public of information on the amount of disclosure made in exchange for the 17-yr patent grant, and hence it invalidates the patent grant.

As a result of the need for its disclosure, an inventor must disclose the best mode of practicing the invention that the inventor knows in drafting a patent application. In some instances, the best mode may be the very commercial product developed by the inventor. However, in other instances the best mode may be an article, machine, or process which is economically or commercially impractical. Nonetheless, this embodiment needs to be disclosed in the patent.

Drafting the Patent Application

Once the record of invention has been assembled and evaluated, a decision may be made as to whether to move forward and draft a patent application. In drafting the patent application, the inventor may work alone gathering the elements of the disclosure which the inventor deems relevant and material to the invention. However, given the technical and legal complexities of patent application drafting, it is more advisable for the inventor to retain a patent agent or attorney. In order for a patent agent or attorney to represent inventors before the U.S. PTO, these individuals must have a degree in one of the sciences or in the field of engineering. Further, a patent agent or attorney must have demonstrated a proven competence in understanding the procedures and rules of the U.S. PTO by obtaining admission to practice before this office.

In drafting a patent application, proceeding methodically through the several steps necessary to produce the type of disclosure legally and technically sufficient to satisfy the requirements of the laws of the United States is absolutely essential to a successful granting of the patent. A first step is to outline those elements of the invention which are absolutely essential to its practice. A body of disclosure should be outlined for each of the essential elements of the claim. This disclosure should describe each element in terms of its function, as well as the parameters that are relevant to the essential nature of the individual element. For example, if a chemical mixture has a component which acts so as to thicken the mixture, it is appropriate to outline the family of constituents that can serve this function. At the same time, a full outline of the disclosure of this individual element will include mention of those chemicals that are preferred for use within the mixture so as to perform the desired thickening function.

Once this process has been completed for each of the essential elements, patent claims may be drafted which cover the invention. These claims will cover, in the broadest sense, only those elements of the invention which are essential. Narrower, more focused claims, however, should also be included within the patent application. These claims may focus on aspects of the invention that the applicant believes are preferred, or may otherwise represent essential aspects of any commercial product that will stem from the invention. Finally, claims should also be drafted to cover alternative forms of the invention. Such alternative forms of the invention may not necessarily be considered to be preferred commercially, but they may present an area where a competitor could attempt to engineer "around" the invention.

The current regulations of the U.S. PTO allow for a total of 20 patent claims with the payment of a minimum fee. Providing claims of varying breadth and scope through the enumeration of essential elements, optional elements, as well as parameters critical to the practice of the invention is desirable. Providing claims of varying scope helps increase the value of any patent by strengthening its validity, making it more enforceable against any infringers, and making it more commercially valuable by enabling the coverage of alternative products and offering the potential of licensing.

Once the claims have been written, a fuller disclosure of the invention may be drafted. This description of the invention will generally follow the outlines of the essential and optional elements. Such an outline will include a functional description of elements including relevant broad and preferred parameters for each of the elements. The description of the invention also should explain the intended interrelationship of the elements that is needed to produce the invention.

Other known embodiments of the invention should also be disclosed to the extent practical. These embodiments can prevent future patenting by third parties if they are published in the applicant's issued patent.

The patent application should also provide a thorough description of the benefits and advantages of the invention and the manner in which it advances to technology. This may be done in a two-part, two-step analysis. The first step is to outline prior developments and inventions in the "Background of the Invention" section of the patent application (Fig. 1b at N). The second step is to describe the advances or benefits of the invention in the "Summary of the Invention" section (Fig. 1b at O). This logical problem—solution format addresses those problems which have been left unmet by the prior inventions. An applicant may also wish to include certain working examples which exhibit the various benefits of the invention. Working examples may also be effective in distinguishing the application from previous inventions as well as in illustrating solutions to problems posed.

With GATT, the U.S. PTO began accepting "provisional" applications as of June 8, 1995. The provisional application provides an applicant the opportunity to gain an early U.S. filing date for a relatively low filing fee without commencing the patent term. Design applications are not included in the provisional application system. The provisional application is not examined by the U.S. PTO except for compliance with formalities, and it has a nonextendable life of one year from the filing date. Drawings must be included if they are necessary for the understanding of the disclosed subject matter.

The provisional application can be filed in a non-English language, but if it is, an English translation is also required. The provisional application must include a cover sheet which (1) confirms provisional status, (2) lists the inventors and the title of the invention, (3) provides the attorney's name and registration number (if applicable), and (4) provides the correspondence address.

To maintain the benefit of the provisional application filing date, a regular utility application must be filed during the pendency of the provisional application, ie, within one year of its filing date, and must include at least one inventor in common with the provisional application. The filing of a

provisional application does, however, commence the one-year Paris Convention priority period. Therefore, foreign filings must be pursued by the first anniversary of the earliest provisional application.

Filing and Examination of the Application

Once the application has been finalized, it should be reviewed by all inventors to make sure that it is a complete teaching of the invention and that the level of disclosure satisfies the legal requirements of U.S. law. The inventors then execute an oath or declaration to this effect. Depending on the structure of the organization with which the inventors are affiliated or for which they work, the inventors may in addition have an obligation to assign the rights for the invention as embodied in the patent application to their employer. In such cases, it is usually appropriate to secure the execution of an assignment by the inventors.

Once the patent application has been reviewed and all formal documents executed, all paperwork including the application is filed with the U.S. Patent and Trademark Office (PTO). Legal regulations govern how a patent application should be filed, and filing is not a simple matter. Further, the correct and appropriate filing of a patent application is essential to obtaining a filing date, which is important to the examination of the U.S. patent application, as well as to the filing of any counterpart applications in foreign countries based on the initial U.S. application.

Once it has been filed, the patent application enters the domain of the U.S. PTO, which is organized by technical discipline into various groups, eg, polymer chemistry, biotechnology, inorganic solid chemistry, as well as organic chemistry. Within each group are specific art units handling areas of technology which are even further focused on specific advances and developments within their respective technical fields.

Figure 3 depicts a generic step-by-step process of examination as it generally occurs within the U.S. PTO. Optional steps are those which may not occur during the process of examining the patent application. Steps designated "if necessary" are those which may not be pursued given the normal course of prosecution. The timing of examination varies depending on the number of patent applications which each group is examining at the time any given application is filed.

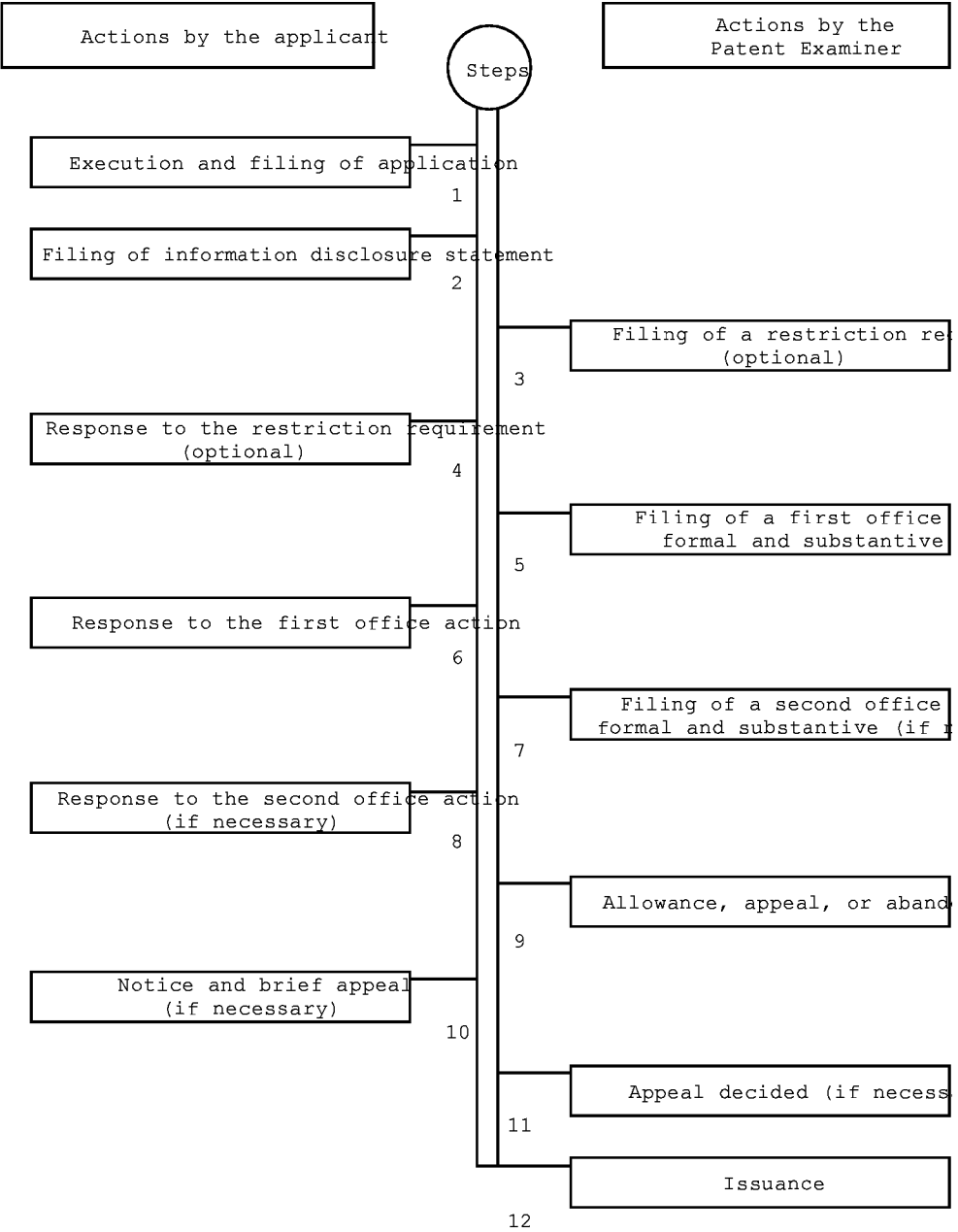


Fig. 3. Timeline for examination in the United States Patent Office.

After the filing of the patent application, the applicant generally files the Information Disclosure Statement (step 2), in which the applicant is required by regulation to list all patents, publications, literature, as well as facts and events which may pertain to the invention disclosed and claimed in the given application. This Information Disclosure Statement is reviewed and considered by the patent examiner in preparation for examining the patent application. The examiner then reviews the claims in the patent application to determine the number and character of inventions disclosed in the single patent application (step 3). The examiner may issue to the applicant a restriction requirement requesting that the applicant select one invention to have examined at that time. For example, if the patent applicant has filed claims directed to a composition, as well as a method of preparing that composition, the examiner may deem that the application comprises more than one invention; the examiner is only obligated to examine one invention. If the examiner requests the applicant to select which invention should be examined first, the unselected invention may be filed at a later date in a subsequent application, called a "divisional application," without loss of right or filing date, as long as it is filed while the first application is pending, ie, before it issues as a patent or is abandoned. Both the restriction requirement as well as the response to the restriction requirement (step 4) are labeled optional, since they may not

actually arise during the course of prosecution.

The first substantive action (step 5) on the merits of the application may occur any time from six to twelve months after filing. This action, generated by the patent examiner and called a first office action, results from the examiner's review of the patent application to ensure its compliance with the formal regulations of the U.S. PTO. These regulations govern definiteness, sufficiency of disclosure, and adequacy of description. In addition, the examiner will have reviewed the applicant's Information Disclosure Statement and conducted a search of prior patents and publications to determine whether the invention was previously known to those in the public. Any publication or patent which has a date preceding the filing date of the patent application being examined may be used against the application as a basis for rejecting the patentability of the applicant's patent claims. The applicant's patent claims will be rejected for lack of novelty if the examiner has found each and every element of those claims within a single publication or patent that has a date preceding the filing date of the application. The examiner will reject the applicant's patent claims as obvious if more than one reference in combination provides an unequivocal disclosure of the claimed invention.

In order to overcome rejections based on prior publications or patents, the applicant often must amend the patent claims to include aspects of the invention which are not found in the publications cited by the examiner as a basis for the rejections. The applicant may also wish to provide properties, characteristics, or advantages of the invention which are unexpected in view of these publications and patents.

In response to the first office action, the applicant may file a series of amendments (step 6) and should provide substantial reasoning and analysis to explain the reasons that the publication(s) cited by the examiner do(es) not disclose the invention as it has been claimed. The patent applicant's response should also comply with the examiner's request for correcting formal problems in the application.

If the examiner believes that all problems or issues have otherwise been resolved in the pending application, the examiner may pass the application onto allowance (step 9). However, if problems still exist with the application the examiner may file a second office action against the pending application, usually a "final rejection" (step 7). At that time, if the patent application is finally rejected, the applicant has a limited opportunity to respond to the examiner. The applicant's second response to the examiner must overcome the outstanding rejection, and provide a response (step 8). Generally, final office actions place the patent applicant in a procedural phase of the examination where the patent application must either be allowed, abandoned, or placed on appeal before the end of the time period set for response to the final office action (step 9). Pursuing an appeal of the examiner's decision involves providing the examiner with notice of the appeal as well as writing and submitting written comments which explain the examiner's error (step 10). The appeal is then decided by an impartial board of administrative judges (step 11).

If the patent application is allowed based on an applicant's response to the second office action, examination is ended. However, if the patent examiner advises the patent applicant that the rejections will be maintained and the applicant views these rejections as insurmountable, the patent applicant may choose to abandon the patent application. If the patent examiner maintains the earlier posed rejections, and the patent applicant disagrees with the examiner, the patent applicant may appeal the examiner's decision to the Board of Patent Appeals and Interferences, which is comprised of administrative judges. The appeal process involves the noticing and briefing of the appeal, and oral argument before and a subsequent decision from the Board of Appeals and Interferences (step 10). Usually the oral argument is presented to, and subsequent decision is received from, a panel of three administrative judges selected from the full complement of the Board. If the Board panel decides in the applicant's favor (step 11), the patent application will be passed to issuance (step 12). If the Board decides in the examiner's favor, the patent applicant will have to consider whether to refile the application and request another round of examination or seek court review.

The mailing of a Notice of Allowance, whether resulting from the decision of the examiner or from a decision by the Board of Appeals, effectively ends the examination of the application. In this instance, the applicant then is required to pay an issue fee. Once the issue fee has been paid, the patent will issue within months.

Starting June 8, 1995, the term of a U.S. patent changed. Design patents retain a 14-yr term. Issued and enforceable patents and patent applications (including continuations and divisionals) in existence before June 8, 1995 have a term which is the longer of 17 years from issuance or 20 years from the original filing date. Patent applications (including continuations and divisionals) filed on or after June 8, 1995 which result in the issuance of a patent have a term of 20 years from filing. This date is measured from the original application filing date. The original application filing date may be the earliest filing date stemming from a United States or Patent Cooperation Treaty (PCT) filing.

Post-Issuance Concerns

Issuance of a United States patent transforms a patent applicant into a patentee, and new concerns may arise relevant to management. For example, the patent should be reviewed to determine formal and substantive correctness. An audit should be taken regularly to determine whether there is a continuing justification to pay the maintenance fees required to keep the patent in force during its effective period. The patentee or patent assignee may have to address concerns of patent infringement or patent validity.

CORRECTION OF ERRORS IN ISSUED PATENTS

A patentee should review the issued patent to ensure that the patent grant is free of errors and contains the intended claims. Errors may arise in a patent application or issued patent during the writing of the patent application, examination of the patent application, or the printing of the issued patent document. The errors may be inconsequential, stemming from misspellings, misprintings, as well as insertions or deletions of text. These errors may have occurred through sections taken by the applicant or the U.S. PTO in transforming the patent application to a printed patent document. The issued patent should also be reviewed for compliance with the formal and substantive requirements of United States law and the regulations of the U.S. PTO.

Other concerns which may necessitate a review of a patent after issuance include the discovery of prior art which was not uncovered during the examination of the patent application. A determination should be made as to whether or not the claims in the issued patent are too broad when viewed in light of this prior art. It may also be the case that someone who participated in the examination of the patent application discovered prior patents, literature, or activities which they knew of but did not cite to the patent examiner. In such an instance, this prior art must also be reviewed in light of the patent claims to determine whether the claims are too broad.

If, upon review of the patent, the patentee discovers that the claims contain a formal error, are too narrow, or are too broad in view of the prior art, the patentee may ask the U.S. PTO to correct this error. There are four administrative vehicles for correcting errors in issued patents. The application of each of these mechanisms is dependent on the nature and severity of the error, as well as the source of its creation.

The Notice of Errors. The first mechanism for correction of errors is called a "Notice of Errors." This document may be filed by the patentee after issuance of the patent with the U.S. PTO and references the patent number, issue date, and the errors contained in the patent. The purpose of a Notice of Errors is to clarify the examination history of the patent and such notice dispositively corrects any misspellings, or typographical errors or omissions. One example of a problem which may be clarified by a Notice of Errors is an omitted chemical bond in a compound used in an exemplary embodiment of the invention. In short, the error is obvious and easily corrected.

The Notice of Errors should resolve those problems which are evident on the face of the patent but which also may be, by their nature, obvious and correctable problems to someone reading only the patent. The Notice of Errors does not result in a further publication by the U.S. PTO, but rather it is instead placed into the examination history of the issued patent and thus is available to anyone who may wish to read this examination history. The Notice of Errors is appropriate for correcting simple matters which do not affect the claim scope or the validity of the patent.

The Certificate of Correction. Another mechanism for correcting the patent is the "Certificate of Correction," which is essentially a petition filed by the patentee to correct minor errors in the patent produced either by the U.S. PTO or inadvertently by the applicant. Unlike the Notice of Errors, a Certificate of Correction does result in an additional publication from the U.S. PTO, and anyone requesting a copy of a patent in which a Certificate of Correction has been filed will also obtain the Certificate of Correction. A Certificate of Correction reflects amendments made during the examination of the patent which were entered by the examiner but not found within the issued patent. The omission of such amendments can be in the body of the patent or in the patent claims. The Certificate of Correction may also be used to correct errors in the issued patent which were not present in the patent application when it was filed. If the error was caused by the patent applicant prior to or during examination, the patent applicant bears the cost of filing and

processing the Certificate of Correction. If the U.S. PTO created the errors during examination or printing, the U.S. PTO bears the cost of producing the Certificate of Correction. The Certificate of Correction cannot be used to amend the patent after issuance. For example, if a review of events surrounding the patent turns up additional prior art that, in turn, requires one or more amendments to the patent, other mechanisms must be used to correct these problems.

Patent Reissue and Reexamination. Reissue proceedings and reexaminations proceedings require the resubmission of the issued patent to the U.S. PTO and should be expedited by the U.S. PTO. However, each requires the additional expenditure of substantial funds and a loss of time in the active life of the patent.

A reissue may be ordered to correct any minor or major mistake which occurred during prosecution of a patent, but the mistake must be one that makes the patent partially or wholly inoperable. Inoperable essentially means that the patent cannot be enforced. For instance, a reissue proceeding can be used to correct inventorship or even broaden claims if the patent is less than two years old. However, such a request to broaden claims in the context of reissue may not be undertaken to recover subject matter canceled during examination. Further, a reissue proceeding may be undertaken to correct formal problems or address newly discovered prior art which affects the scope of the claims. The nature of a reissue proceeding directs that this mechanism should be used only when the validity of the patent is in question owing to the error or problem in question.

A request for reexamination may be made by the patentee, a third party, or the Commissioner of the Patent and Trademark Office whenever a new question of patentability arises. This new question of patentability has to be raised in the form of a publication such as a journal article or a patent which was not considered during the prior prosecution. Reexamination is a more expedited and economical means of receiving a judgment on whether or not a patent is valid in advance of litigation.

By requirement, the patent generally must be resubmitted to the U.S. PTO for reexamination. If the examiner and the patentee differ as to their findings concerning prior art or the ultimate scope of the claims, reexamination can take an extended period of months if not years to complete, all at substantial cost to the patentee. In addition, reexamination and reissue proceedings allow for varying levels of participation by third parties. As a result, before undertaking any such proceedings a patentee should consult legal counsel to define a legal strategy and choose an appropriate forum for correction of the patent in question.

PATENT MAINTENANCE FEES

On the date a patent issues, it has a 20-yr life measured from the earliest filing date relied on for priority. Under current U.S. law, the patentee is required to pay maintenance fees, a policy stemming from an interest in the public in practicing the technology covered or claimed in the patent. After a patent issues, the claims are generally very important commercially and provide the patentee with relatively easily exercised rights to prevent others from making, using, or selling that which is found in the claims. However, as the patent grows older, the public interest in practicing the technology grows stronger. Often the claims become less important commercially and the commercial value of the claims then needs to be assessed in view of the expense of maintaining the patent.

Payment of maintenance fees is required at the fourth-, eighth-, and twelfth-year anniversaries of the date of issuance of the patent. The costs of these maintenance fees vary from year to year depending on the regulations of the U.S. PTO. The first maintenance fee tends to be fairly slight, allowing for the further enforcement of the patent with little economic burden on the patentee. The second maintenance fee tends to increase the payment from the first maintenance fee by a factor of two. The third maintenance fee tends to be substantial, increasing the payment from the first maintenance fee by a factor of three.

The patentee should develop and implement a policy for auditing its patent portfolio in the process of paying maintenance fees to the U.S. PTO. This practice should also be used to justify the further payment of annuities to foreign national patent offices. Maintenance fees and annuities can constitute a substantial portion of funds expended in the protection of patents over a year's time. Further, without a tangible, real commercial value or advantage stemming from the patent, there may be little justification for maintaining the patent over its last five years of life.

Factors which should be considered in auditing or otherwise determining continued justification for maintaining a patent include the following: (1) Do the patent claims cover a product made by the patentee? If so, what is the level of income provided? (2) Do the patent claims cover a product made by an infringer or a product which is likely to be made by an infringer within the remaining term of the patent? What is the potential gain in income for the infringer, and what is the potential royalty? (3) Is there a potential to sell or to license the patent? (4) Is there any other commercial justification for maintaining the patent through payment of the maintenance fee?

PATENT INTERFERENCE

An interference is a contested action in the U.S. PTO to determine inventorship between two or more patent applicants or between at least one patentee and one or more patent applicants. The principal contest in an interference concerns the right to claim the invention. The interference action results from U.S. law, which awards patents to the first inventor, generally irrespective of patent application filing date. In the simplest situation, an interference occurs when a pending application discloses and claims the same invention which is claimed in at least one other copending application or issued patent.

The interference proceeding is declared by the patent examiner and occurs in the U.S. PTO. Once an interference is declared, a determination is made as to the exact subject matter constituting the invention of the interference and who filed the first patent application on that invention. The first applicant becomes the senior party to the interference. The junior party has the burden of proving that it was prior in time as to its date of invention.

A patentee or patent applicant may win an interference proceeding by proving the right to the invention as the first inventor. Alternatively, a patentee or patent applicant may win an interference proceeding by default. If the invention was known to the public prior to the first date of invention, none of the parties to the interference have a right to claim the invention.

LEGAL ACTIONS BASED ON PATENTS

The issuance of a patent initiates a term during which the patentee may enforce its rights, ie, the patentee may prevent others from making, using, or selling that which the patent claims. To literally infringe upon the patentee's rights, another person, business, or organization must make, use, or sell something which has each and every element found within the claims of the patent in question. Patent infringement may also occur if the action in question contributes to or otherwise induces the making, using, or selling of something which contains each and every element of that which is found in the claims of the patent.

An action for patent infringement may be based on one or more claims of the patent. The patentee may also bring an action for infringement of patent rights if it believes that the actions of a third party are equivalent to that which is literally recited in the patent claims. If infringement of patent rights is found by court of law, the patentee may receive remedies that include monetary damages, attorney's fees, and injunctive relief.

One accused of infringing on patent rights may defend against the action by showing that they have not made, used, or sold something which includes each and every element found in a claim of an issued patent. Further, one may also defend against a legal action for infringement of patent rights by clearly showing that the patent in question is invalid, ie, that it lacks novelty, is obvious, has not complied with the formal disclosure requirements of the U.S. PTO, or does not designate the proper inventor.

Legal actions based on patents almost always have tremendous commercial significance to the parties involved. The factual and legal issues surrounding and relating to these legal actions can be complex, burdensome, and not easily resolved. Any number of oral or written opinions are necessary from legal counsel which provide the appropriate advice or defenses. Further, the guidance provided by statute and legal precedent may often be lacking. A thorough discussion of legal actions relating to patents is beyond the scope of this article. One considering, or threatened by, a legal action involving a patent should retain competent legal counsel.

Foreign Patent Prosecution

The foreign filing of a patent application is an immensely complex task, requiring retention of foreign patent lawyers or patent agents, complying with highly specific rules of foreign practice, and usually requiring a significant expenditure of capital. However foreign patents can provide significant commercial opportunities in valuable international markets. Further, various systems for obtaining patent protection, put in place by multinational treaties, are allowing most organizations to operate on a commercial level which is not national or regional, but global.

For many years, the method of obtaining foreign patent protection corresponding to a U.S. patent application was to file separate, individual patent applications in selected foreign countries. Each of the applications had to be written to conform with the national requirements of the country in which it was filed.

The national laws of most countries are unique to the particular country. However, most industrialized countries are parties to one or more International Conventions which provide for the filing of foreign patent applications. For example, pursuant to the Paris Convention, if an application is filed in a second country within one year after filing the application in the first, "home" country, and if certain legal formalities are met, most foreign countries will treat the foreign-filed application under its own laws as if it had been filed on precisely the same date as in the original, home country application.

The benefit of these treaties generally relates only to obtaining a retroactive filing date, and the individual laws of each foreign country still apply. However, this rule has some important consequences. For example, while the United States permits a one year grace period in which to file a patent application following any public use or sale or other disabling act, many foreign countries do not. Consequently, even with the foreign filing treaty benefits, it has become necessary to file U.S. applications before there has been any disabling act, even though it is not required by U.S. laws.

The Patent Cooperation Treaty. A number of foreign countries have signed a Patent Cooperation Treaty (sometimes known as the PCT). Filing a single application in the English language under this treaty allows an applicant to designate any of the countries which are signatories of the PCT. The treaty establishes a system for the centralized filing and searching of a single national application and operates to establish a common filing date for all designated countries. Thus, the initial application, when accompanied by the proper designation of countries and payment of applicable fees, is deemed to be the equivalent of foreign filings on a country-by-country basis of the same application in each country designated by the applicant.

Under this filing procedure, once the single International PCT application is filed a patentability search is conducted by an approved search office. After the search is completed, the search report is published and reported to the applicant.

If the applicant decides to proceed, the PCT patent application will publish. At this time, the applicant may either proceed to file the PCT patent application in each designated national country or request examination of the PCT application within the framework of the PCT. If the applicant requests examination of the patent application within the PCT, an Examiner will file a written action. The applicant may then amend the application by resubmitting new claims and new application pages.

These amendments to the application can be made in English. The Examiner reviews the applicant's submissions and issues a final written opinion on the patentability of the application. At this time, the applicant must pursue national filings within those national and regional patent offices in which the applicant intends to obtain protection. Copies of the PCT patent application and written opinion are then distributed to the National Patent Office by the World International Patent Office (WIPO).

Publication of the complete application occurs as a single International Patent Application 18 months after the first, home country filing date. In theory, this is a simpler procedure for the initial filing of multiple applications in individual countries and avoids a duplication of search efforts. Further, by postponing the requirements for filing translations of an application until after the search results are known, a potential savings is afforded to an applicant who, after reviewing the search results, decides not to continue prosecution of the application in one or more countries which have been initially designated. However, some of these cost savings are offset by special fees, eg, country designation fees, applicable to any PCT application. Also, the process of obtaining a patent in any particular country may be delayed for quite some time relative to the length of time that one would expect if a national application had been filed directly in that particular country. Thus, PCT examination permits delaying the decision as to which particular country the application should be filed.

In the PCT examination process, national patent applications may be filed at the end of Chapter I (18 months from the home application filing date), or at the end of Chapter II (30 months from the home application filing date). If the examination of the U.S. home application ends successfully before completion of the examination of the counterpart application in PCT Chapter I, the PCT Applicant may file national applications including the claims examined and allowed by the U.S. PTO without proceeding to PCT Chapter II processing. A brief timeline of patent application examination under the PCT is provided in Table 1.

Table 1. Patent Application Examination by the PCT Process

Date	Action
day 1, month 1	filing of first national (home) application
day 1, month 13	PCT application must be filed by this date
month 16	publication of PCT search report
month 18	publication of PCT application
month 19	applicant must file a demand for examination of PCT application to maintain the application, or complete national filings in the intended countries
month 23	if a demand is filed, examiner issues first examination report
month 26	applicant's response to examination report is due
month 28	final examination report is issued by PCT examiner
month 30	end of PCT examination; patent application enters in national or regional examination

Regional and National Patent Application Filings. Other methods of obtaining patent protection in foreign countries include national filings and filings undertaken under regional conventions, eg, the European Patent Convention (EPC), to which most European countries belong. National filings can be time-consuming, laborious, and expensive. Most often examination is undertaken in the local language of the country. The expense of retaining local associates, securing translations, and paying local fees usually results in higher costs than are incurred under the PCT or certain regional conventions. Still, certain countries, such as Egypt, Hong Kong, Indonesia, Malaysia, Turkey, South Africa, the Philippines, and Argentina, have maintained their independence from the various world and regional conventions.

Under the EPC an applicant may file a patent application in one or all of the European countries that are signatories by filing a single application. Unlike the PCT, the EPC is actually a system of law, common to all of the present member countries, established for the granting of European patents.

An application filed under this treaty results in a single application which is processed by the European Patent Office, located in Munich, Germany. The European Patent Office, like the U.S. PTO, performs the patent function for all of its member states, but only for patent applications filed under the EPC. At the European Patent Office, the application is examined and granted. The initial filing and most prosecution under the EPC can be made in the English language. When the patent is granted, it is actually granted in the form of a number of identical patents, each of which applies to one of the countries which was designated at the time of filing the European patent application. Each of these patents is interpreted and enforced according to the applicable national laws in the country to which the patent applies.

Even in countries which are not signatories to either the PCT or various regional conventions provided around the world, patent application examination generally follows a fairly standard pattern. After the first national, home application is filed, subsequent applications may then be filed in other countries, within the 12-month time period if such a grace period is provided. If this grace period is not provided, the patent application(s) which are to be filed in non-Paris Convention Countries have to be filed before any event occurs that may destroy the novelty of the invention. Further, under U.S.

practice, any invention that is the subject of a U.S. patent application and that is also to be filed outside the United States must be given a foreign filing license by the U.S. PTO prior to the foreign filing.

Once the patent application has been filed in a foreign national or regional patent office, a series of events take place. First, the patent application is assigned to a patent examiner within the regional or national patent office. The examiner generates a patent examination search report, which will be the basis for the examination of the application and which allows the applicant to evaluate the invention in the context of those patents and patent applications which have been previously filed around the world. Unlike the United States, which maintains patent applications in secret until patent issuance, most national and regional patent offices publish patent applications at a regular interval after the first filing date. In many countries this publication occurs at the 18-month anniversary from the first filing date of the application. In some countries, eg, Taiwan, this publication may occur at any time once examination has been successfully completed. Further, in at least one country, Japan, patent applications are published in each stage of examination.

The next step in examination is a request for the examiner to take action or take the application up for examination. Examination in foreign countries can be very complex, rigorous, and formal. Most countries require a substantive examination of patent applications including a review for compliance, with formal requirements in the format of the patent application and patent claim. In addition, most foreign national and regional patent offices require examination of applications for novelty as well as a concept called inventive step. Although the requirement that an invention have an inventive step is somewhat complex, this requirement is analogous to the requirement that inventions not be obvious under U.S. law.

In the process of examination, a foreign national or regional patent examiner may generate any number of office actions, to each which the applicant is required to respond. When dealing with foreign national and regional patent offices, the applicant is often required as a matter of practical necessity to retain an attorney in the foreign country in which patent protection is being sought. Such an attorney may assist in securing translations, providing practical insight into the legal requirements of the national or regional patent office, and in providing practical insight into the interpretation of publications cited against the application.

Once the application has been determined to be acceptable by a foreign national or regional patent examiner, it is generally published as allowed or granted. At this time, the laws of most foreign national or regional countries or patent conventions allow for the opposition of the allowance or grant by any third party who may deem the invention unworthy of a patent grant. If opposed, the patent grant will undergo an additional examination. If unopposed, after the specified time period the application is granted.

After granting, the applicant must comply with annuity requirements, the necessity of commercially exploiting the invention in the foreign country, any requirements to grant compulsory licenses, and it must also undertake the enforcement of its patent rights. Annuities are generally taxes levied on the patent grant on an annual basis. In many countries the annuities are levied on a claim-by-claim basis, so that the more claims a specific patent contains, the more expensive that patent is to maintain over its lifetime. In certain countries, such as those in the Pacific rim, patent rights can become a very expensive asset to maintain for the life of the patent.

Enforcement of foreign patent rights is of concern. In some countries, enforcement proceedings to prevent another party from using a patented invention can be difficult to initiate owing to the expense, the time period required for the enforcement, or the overall practicality of any remedy provided by the laws of the country. As a result, given the expense of securing foreign patent protection, a principal consideration in the decision as to whether to file an application should be whether the applicant will ever be able to enforce any patent rights actually obtained in that country.

In evaluating foreign patent protection for an invention, it is necessary to (1) select the countries where protection is desired; (2) determine which of those countries are participating in the European Patent Convention (or regional treaty), which are participating in the Patent Cooperation Treaty, and which are not participating in either treaty; (3) evaluate the importance of the invention; (4) consider the level of inventiveness or sophistication of the invention; (5) evaluate the need or lack of need for secrecy; (6) consider the present or imminent likelihood of infringement by others; and (7) consider licensing needs.

Trade Secret Rights

An alternative to patent protection for advances, developments, ideas, and applications is to treat such information as a trade secret. The protection of trade secrets relies on the development of ideas, applications, and advances which are not found in the public domain. Trade secrets, by definition, are kept in confidence by their owner, disseminated only to those who accept an absolute obligation of confidentiality, and then only for purposes of which the trade secret owner knows and approves.

If a trade secret is believed to have been violated, a judge must initially decide whether or not it actually existed. Such determination is based in part on the manner in which the trade secret was protected and also on such considerations as the commercial value of the information, the manner in which the information was safeguarded, and the manner in which the information was stolen or otherwise found in the public domain. These are also some of the initial factual determinations which must be made when considering trade secret protection.

In evaluating the application of trade secret protection, a matter of further concern is the commercial relevance of the trade secret. For example, consider whether an improved chemical process for fabricating a semiconductor should be protected by means other than patenting. If it is not possible to ascertain the nature of the new process by close analysis of the semiconductor, the basis on which to file and easily enforce a patent may not be present. By disclosing the process in the form of a published patent the process owner obtains a 17-yr grant of patent rights upon issuance in the United States, but the patent may disclose a key element of the patent owner's business to the patent owner's competitors. Since the process is not evident from the finished semiconductor, the patent owner's competitors would not, but for the issuance of the patent, be able to learn of the process. Also, the patent owner may face the difficulty of not being able to discern whether another party is using the patented process. As a result, the patent owner may have a very difficult time enforcing patent rights against competitors who are using the process in violation of the patent.

The life of a trade secret may extend indefinitely if the owner of the secret has taken the proper steps to safeguard the invention, in contrast to a 17-yr patent term, after which time the invention is in the public domain. Traditionally, trade secrets have been protected by confidentiality agreements, nondisclosure agreements, and employment agreements.

The Creation of a Trade Secret

Because there is no "federal law of trade secrets," protection of trade secrets is often left to the variability of the criminal and civil laws of the 50 states. To the extent that a trade secret is property, violation, theft, or misappropriation of the trade secret may be the subject of criminal penalty. To the extent that a trade secret is bound to rights, violation or misappropriation of the trade secret may be the subject of civil penalty. Significant effort, however, has been made in developing a uniform body of law to apply to ideas and innovations which may be the subject of this form of protection.

Trade secrets may be any type of information, eg, formulae, patterns, compilations, forms, programs, devices, techniques, and processes, as well as any patentable subject matter. However, in order for it to be a trade secret, there must be definite economic value in the information not being known to the public or readily determinable by a third party.

The trade secret must also be the subject of reasonable efforts to maintain its security, though the disclosure of a trade secret does not necessarily end its protectable life. Rather, an evaluation must be made as to whether the disclosure was made by someone who knew, or should have known, that the information was a trade secret. If so, trade secret rights may still be protected.

In determining whether confidential information is indeed a trade secret, knowledge of the information generally, as well as knowledge of the information by those in a given business, must be considered. Other factors include measures taken to safeguard the information; the value of the information within any given business setting; the investment in time and capital made to develop the information; and the ability of others to acquire the information through proper channels.

To summarize, in order to be considered a trade secret, the information (1) must not be generally known or readily ascertainable; (2) must provide a competitive advantage; (3) must have been developed, maintained, or acquired at the trade secret owner's expense; and (4) must be the subject of the trade secret owner's intent and efforts to keep it confidential.

- Vital information a company may wish to protect will invariably be connected to the company's products or services. In order to facilitate protection, a policy or program should be implemented requiring regular recognition of such information by employees who create, use, or otherwise have access to it.
- The protection of a trade secret is a complex task dependent upon any number of factors. The mere formation of an intention to maintain information as secret is not enough; actual safeguards must be put into place. The owner of a trade secret must identify the information as a trade secret and protect the information from disclosure. Means used to prevent disclosure might include the following:
- Guarding entrances to the facility in which the information is kept and checking briefcases, purses, and the like and locking or making unguarded entrances inaccessible.
 - Using employee and visitor identification badges.
 - Maintaining logs of all who visit the facility during the day, denying visitor access to sensitive areas of the facility, providing visitor escorts within the facility, and guarding experimental, developmental, and prototype work from public view.
 - Developing an organizational policy on trade secret information and communicating that policy to employees; promoting a program to identify all commercially important or competitively sensitive information for protection as trade secrets, and further identifying such information by clearly labeling it.
 - Storing and maintaining trade secrets under lock and key in a segregated area that is not accessible to the public.
 - Limiting access to trade secrets to those having an obligation to maintain it as confidential.
 - Using agreements to prevent employees who are exposed to trade secrets from disclosing this information if or when they change employment.
 - Destroying trade secret information by means that will prevent its disclosure, eg, incineration or shredding.
 - Monitoring or providing clearance for employee publications, lectures, and other public activities related to the business of the organization.
 - Monitoring or clearing employee activities which involve removing any business-related information or objects from the facility.
 - Maintaining a user log, either manual or electronic, on all photocopying equipment.
 - Using photographic, electronic, or keyed access and monitoring equipment.
 - Consistently protecting all trade secret information to the same level, including consistently investigating any concerns over the theft or breach of trade secret protection.

Exploitation of Trade Secrets

Trade secrets become unprotectable when they are found in the public domain, are independently developed, or are disclosed out of confidence. Events of the latter type may occur in any number of controlled or uncontrolled situations. For example, a promotional event such as a trade show or a required disclosure to a governmental agency may result in disclosure of the trade secret. Further, publications in journals or magazines which may be necessary to promote products may lead to a disclosure of trade secrets. Idle correspondence, conversations, or communications with sales associates, suppliers, or distributors may also result in disclosure of trade secret information.

A trade secret owner may also beneficially exploit the trade secret through licensing, sales, or various other business ventures based on the confidential information. Such cooperative ventures often raise other issues. Exploitation of trade secret information may also occur through the unintended disclosure of this information to the public. Generally, the people who learn of trade secret information tend to be the trade secret owner's employees, customers, licensees, suppliers, and joint venture partners.

In business transactions the parties should have a clear understanding of exactly what constitutes trade secret information and consider how the information will be used and who will retain ownership rights. If the transaction is a pure and simple sale, concerns over ownership may be meritless. However, such concerns might be well-founded, if further research or commercial development involves similar information. It may also be necessary to consider whether the seller should be allowed to compete against the buyer in ventures involving the same or related information. These are just some of the issues which arise with the sale of the trade secrets.

When licensing or otherwise undertaking a joint venture based on trade secret rights occurs, other considerations arise. For example, research efforts invariably give rise to additional information which may be the subject of trade secret or even patent protection. If this additional information is derived from the licensed or shared body of initial information, consideration should be given to ownership, further protection, eg, who files and pays for patent protection, and at the end of the agreement how, or even if, this information should be divided. Commercial partners of the trade secret owner should not be provided this information except under the strictest obligations of confidentiality. Keeping in mind that a trade secret owner's supplier probably also supplies the owner's competitors, such relationships are often ready conduits for the dissemination of confidential information. Further, license and joint venture agreements regularly contain confidentiality provisions with substantial penalties for any violations that may occur.

In the case of contract and noncontract employees, a rigid program devised for the identification and protection of trade secret information should be implemented. Confidentiality agreements should be signed by all employees at the time of hire. The agreement may contain a number of provisions on the use of information during and after employment. The significance of the program, including the employee's responsibilities and the company's rights, should be clearly explained to the employee. The employee should also be given tools for maintaining information as a trade secret; for example, the simple use of bound notebooks for maintaining laboratory experiments is almost a universally accepted standard practice. The use of a resource person for questions on identification and protection of established and newly developed trade secret information is also a good practice.

Employees should be regularly briefed on the organization's trade secret program. These briefings should be directed toward the clarification of issues, questions, and concerns. If an employee leaves the company, the organization should remind the employee of its rights and that the obligations of confidentiality continue to bind the employee even after termination of employment.

In short, trade secret information should be disseminated only when commercially necessary, only under obligations of strict confidentiality, and only with definite penalty provisions for improper use or further dissemination.

Even so, a trade secret owner may wish to outline a plan for further protecting the trade secret information. For example, in instances where publication of a trade secret is necessary for commercial exploitation, the filing of a patent application may be an adequate substitute for the complete dedication of rights to the public. If the information would satisfy the requirements of U.S. patent law, then, despite perceived difficulties in enforcement of any patent rights obtained, the best defense against theft or unauthorized use may be obtaining patent rights covering this information.

Violation of Trade Secrets

Trade secret rights are generally violated through an unauthorized use by someone other than the owner. This use may take the form of theft or misappropriation for later use in a commercial product. The unauthorized use can also take the form of an unauthorized disclosure to a third party who is not bound to keep the information confidential.

Another form of misappropriation is the disclosure or use of a trade secret of another without consent, by a person who used "improper means" to acquire knowledge of the trade secret. "Improper means" generally include theft, bribery, misrepresentation, breach or inducement of a breach of a duty to maintain secrecy, or espionage.

The finding of misappropriation is highly dependent upon the protection that the owner has given the trade secret, as well as the notice provided to employees. That is, an employee has a duty of confidentiality to his employer for that information which is considered secret, but the employer must provide the employee fair notice of the confidential nature of the material.

An additional concern for trade secret owners is that like other legal actions, there is a definite limitation to the time period for bringing an action for misappropriation of trade secrets. As a general rule, such a legal action must be brought within three years after the misappropriation is discovered. Remedies for trade secret misappropriation can include injunctive relief and money damages, as well as attorneys' fees for bringing the action.

BIBLIOGRAPHY

"Patent Practice and Management" in *ECT* 2nd ed., Vol. 14, pp. 552–583, by R. Calvert, Consulting Patent Attorney; "Patents—Practice and Management" in *ECT* 3rd ed., Vol. 16, pp. 851–889, by J. B. Gambrell, C. M. Cox, and P. K. Krieger, Pravel, Gambrell, Hewitt, Kirk, Kimball & Dodge.

1. **U.S. Pat. 5,157,128** (Oct. 20, 1992), S. E. Hagen and M. J. Suto (to Warner Lambert Co.).
2. **U.S. Pat. 5,241,990** (Sept. 7, 1993), J. D. Cook (to Inlet Medical, Inc.).
3. **U.S. Pat. 5,020,452** (June 4, 1991), D. T. Rybak (to Murya, Inc.).
4. **U.S. Pat. 5,112,462** (May 12, 1992), R. L. Swisher (to Sheldahl Inc.).
5. **U.S. Pat. 4,997,573** (Mar. 5, 1991), G. R. Browne (to Golconda Engineering and Mining Services, Pty. Ltd.).
6. **U.S. Pat. Des. 334,420** (Mar. 30, 1993), E. J. Gladfelter, T. O. Outlaw, J. L. Copeland, R. K. Schulz, D. K. Boche, and J. W. Peterson (to Ecolab Inc.).
7. **U.S. Plant Pat. 3,360** (June 19, 1973), D. L. Armstong.

General References

M. Arther Auslander, *Chemical Inventions and Chemical Patents*, Clark Boardman Company Ltd., New York, 1964.

J. L. White, *Chemical Patent Practice*, Patent Resources Group, Inc., Apr. 1993.

H. L. Hanson, *Creativity, Innovation and Intellectual Asset Management*, Honeywell, Inc., 1984.

A. M. Hale, *Patenting Manual*, 2nd ed., S-P Inc., 1993.

Code of Federal Regulations, Title 37, Patents, Trademarks, and Copyrights, U.S. Government Printing Office, Washington, D.C., 1985–1995.

Colloquim on the Patent Cooperation Treaty, World Intellectual Property Organization, 1993.

Commissioner of Patents and Trademarks—PTO Annual Report 1992, U.S. Dept. of Commerce, Washington, D.C., 1993.

A. W. Deller, *Deller's Walkers on Patents*, The Lawyer's Co-Operative Publishing Co., 1972.

H. R. Mayers and B. G. Brunsvold, *Drafting Patent License Agreements*, 3rd ed., Bureau of National Affairs, Washington, D.C., 1990.

General Information Concerning Patents, U.S. Department of Commerce, Washington, D.C., Apr. 1989.

United States Patent and Trademark Office, *Guide for the Preparation of Patent Drawings*, U.S. Government Printing Office, Washington, D.C., 1993.

R. Goldscheider, T. Arnold, and W. Poms, *Law and Business of Licensing: Licensing in the 1980s*, Clark Boardman Company Ltd., New York, 1989.

J. Pagenberg and B. Geissler, *License Agreements: Patents, Utility Models, Know-How, Computer Software*, Cad Heymanns, 1989.

J. L. Landis, *Mechanics of Patent Claim Drafting* 2nd ed., Practising Law Institute, 1974.

T. Arnold, *Patent Alternative Dispute Resolution Handbook*, Clark Boardman Co., Ltd., New York, 1991.

S. A. Becker, *Patent Applications Handbook*, Clark Boardman Co., Ltd., New York, 1992.

E. Ridsdale, *Patent Assignments*, Baker, Voorhis and Co., Inc., 1955.

World Intellectual Property Organization, *Patent Cooperation Treaty (PCT) and Regulations Under the PCT (As in Force from July 1, 1992)*, WIPO, 1992.

G. K. Koenig, *Patent Invalidity*, Clark Boardman Company Ltd., New York, 1980.

Southwestern Legal Foundation, *Patent Law Annual*, Matthew Bender and Co., 1974–1990.

D. S. Chisum, *Patent Law Digest—1993*, Matthew Bender and Co., 1993.

I. M. Aisenberg, *Patent Law Precedent: Key Terms and Concepts*, 2nd Edition, Little, Brown and Co., Boston, Mass., 1991.

T. Arnold and L. H. Pretty, *Patent Litigation 1993*, Practising Law Institute, 1993.

L. Horowitz, *Patent Office Rules and Practice*, Matthew Bender and Co., 1992.

I. Kayton, *Patent Practice*, 6th ed., Patent Resources Institute, 1995.

J. M. Samuels, *Patent, Trademark, and Copyright Laws*, 1993 ed., Bureau of National Affairs, Washington, D.C., 1993.

E. B. Lipscomb, III, *Walker on Patents* 3rd ed., The Lawyers Co-Operative Publishing Co., 1984.

J. W. Baxter, *World Patent Law and Practice*, Matthew Bender and Co., 1991.

Guidelines for Examination in the European Patent Office, European Patent Office, Munich, Germany, 1985.

J. G. Sheldon, *How to Write a Patent Application*, Practising Law Institute, 1992.

J. Rosenstock, *Law of Chemical and Pharmaceutical Invention: Patent and Nonpatent Protection*, Little, Brown and Co., Boston, Mass., 1993.

U.S. Department of Commerce, *Manual of Classification*, U.S. Government Printing Office, Washington, D.C., Dec. 31, 1991.

U.S. Department of Commerce, *Manual of Patent Examining Procedure*, U.S. Government Printing Office, Washington, D.C., 1983.

J. E. Hawes, *Patent Application Practice* 2nd ed., Clark Boardman Callaghan, New York, 1993.

P. D. Rosenberg, *Patent Law Basics*, Clark Boardman Callaghan, New York, 1992.

P. D. Rosenberg, *Patent Law Fundamentals*, Clark Boardman Callaghan, New York, 1988.

B. Kramer and A. D. Brufsky, *Patent Law Practise Forms: Rules, Annotations, and Commentary*, Clark Boardman Co., Ltd., New York, 1985.

J. W. Schlicher, *Patent Law: Legal and Economic Principles*, Clark Boardman Callaghan, New York, 1992.

L. Horowitz, *Patent Office Rules and Practice Forms*, Matthew Bender and Co., 1984.

D. A. Chisum, *Patents*, Matthew Bender and Co., 1994.

A. Jacobs, *Patents Throughout the World*, Clark Boardman and Co., Ltd., New York, 1988.

PCT Applicant's Guide, WIPO, 1991 ed.

R. Goldscheider, *Technology Management*, Clark Boardman Co., Ltd., New York, 1988.

M. F. Jager, *Trade Secrets Law*, Clark Boardman Co., Ltd., New York, 1988.

S. H. Liberstein, *Who Owns What is in Your Head?; Trade Secrets and the Mobile Employee*, Hawthorn Books, 1979.

R. M. Milgrim, *Milgrim on Trade Secrets*, Matthew Bender and Co., 1994.

Burchfiel and co-workers, in I. Kayton, ed., *Practice Under Gatt's Uruguay Round Agreements Act*, Patent Resources Institute, 1995.

John J. Gresens
Merchant & Gould

PATENTS, LITERATURE

- Background information,
- Secondary sources of patent information,
- Advances in patent documentation,
- Types of patent information searches,
- On-line database searching methods,
- Cross-file and multifile techniques,
- Patent databases,
- Archiving and document delivery,
- Other technological initiatives,

Background Information

History. Patent systems were established by industrialized countries during the Industrial Revolution of the eighteenth and nineteenth centuries to provide incentive for the development of technology and information to later inventors about the advances made by innovators. Patent rights are limited to control of the manufacture and use and sale of the invention claimed in the patent; a patentee has the right to practice the claimed invention only to the extent that this does not require practicing inventions claimed in patents owned by others. The patentee has the right to license, reassign, or sell the rights conferred by the patent and to sue for infringement, unauthorized manufacture, use, or sale of a product, a process, or an apparatus covered by the patent claims (see PATENTS AND TRADE SECRETS). Patents are granted by national governments and have effect only within the granting state. Patent rights and collections of patent literature documenting these rights have long existed in each industrialized country. The internationalization of commerce in the late nineteenth century necessitated the filing of patent applications in each country where the innovator wished to exclude others from practicing the invention, providing motivation for the founding of an international treaty establishing standardized treatment of patent applicants. Since 1883, the Paris Convention for the Protection of Industrial Property has guaranteed that an applicant for a patent in one member state may file applications for patents in all other member states within one year of the original filing date and will be given rights to the claimed invention as of the priority date established by the first filing.

The second half of the twentieth century has witnessed a sharp increase of activity in research and development, as well as an increased internationalization of technology-based industries. As a result, there have been significant changes in the patent literature, the chief literature of technology. The number of countries publishing patent documents has increased as former communist and Third World countries have enacted patent laws. The number of patent-issuing authorities is also growing rapidly as a result not only of the emergence of new nations formerly embedded within the Soviet Union and other communist countries of central and eastern Europe, but also of the enactment of new patent laws by other countries in response to the intellectual property provisions of the General Agreement on Tariffs and Trade (GATT), which established the World Trade Organization (WTO), and to the North American Free Trade Agreement (NAFTA). The ideal of full harmonization of patent laws among countries has often been discussed, but seems far from being realized; nevertheless, significant changes have been made in the patent laws and procedures of individual countries.

The changes have been in patent documents themselves as well as in the means of documentation. The need to cope with a rapidly increasing volume of patent applications led a number of patent offices to switch, mostly between 1964 and 1979, from a system in which all applications were examined and only those found to be worthy were issued a patent, to a system in which all applications are published and may or may not be examined at a later date or will ever become a patent. In the 1940s most published patent documents were patents; in the 1990s most are unexamined applications.

Until the 1970s, all patents were effective only for the individual countries by which they were issued. In the 1990s there are several types of international patent documents, eg, the European patent, which provides patent rights granted centrally by the European Patent Organization (EPO), and which is enforceable after national registration in as many as 17 European countries at this writing. A Eurasian Patent Convention was established in 1994 by 11 former member states of the Soviet Union. Two regional organizations cover a number of African nations: the African Intellectual Property Organization (OAPI) for 14 French-speaking countries and the African Regional Industrial Property Organization (ARIPO) for 11 English-speaking countries. Patent Cooperation Treaty (PCT) applications provide a means for filing applications for patents in multiple patenting authorities via a single application. PCT applications receive a preliminary examination and are published by the World Intellectual Property Organization (WIPO) prior to examination under the national patent laws of each of the member states (78 in 1995) designated by the applicant. Rights under granted European patents can at present be extended to two additional PCT contracting states, Lithuania and Slovenia; it is expected that additional PCT states will be added to this list. Change in this field has been rapid in the 1990s. Table 1 shows some of the milestones in the development of the primary patent literature during the last third of the twentieth century.

Table 1. Recent Milestones in the Development of Primary Patent Literature

Year	Country or authority	Significant development
1964	the Netherlands	first principal examining office to switch to universal publication and deferred examination switch to universal publication; huge backlog of pending cases published, often at rate of over 1000 wk, which strained documentation services
1968	FRG	
1971	Japan	switch to universal publication, output rose quickly above 100,000 yr (~350,000 yr in 1992–1994); language and numbers make quality documen-tation a substantial problem
1979	European Patent Office (EPO)	single patent covering multiple countries; tended to supplant national patent offices; increased share of English-language publications
1979	World Intellectual Property Organization (WIPO)	single application submitted to multiple countries and regional offices; further increased share of English-language documents; has become very significant in 1990s
1980	United States	periodic maintenance payments required for granted patents having file dates later than Dec. 12, 1980
1995	United States	switch to 20-year term from file date; may begin publishing unexamined applications

Patent laws provide for several stages in the life of an application for a patent on an invention. The pattern followed by patent laws in effect in most industrialized countries during the nineteenth and early twentieth centuries, and still in effect in the United States in 1995, calls for the examination of all patent applications to certify that the claimed invention meets the national standards for novelty, usefulness, and inventiveness. The owner of the technology to be patented files application papers that include a specification containing a description of the invention to be patented (called the disclosure) and claims defining the limits of the invention to be protected by the patent, a formal request for the issuance of a patent, and fees. Drawings of devices and apparatuses, electrical circuits, flow charts, etc., are an important part of the disclosures of most nonchemical and many chemical patents.

The national patent-issuing authority assigns an application serial number, examines the application papers to see that all requirements are met, examines the claims of the application to determine whether patent rights are justified in view of the earlier disclosures in the prior art, and corresponds with the applicant to negotiate any amendments that might be required. When a suitable scope for the claimed invention is agreed to, the patent is issued and the patent specification is published. When the patent examiner determines that no patentable invention has been claimed, the patent application is abandoned. In the United States, unexamined applications were not published under the law that prevailed until 1995, thus the patent office provided no direct evidence that such applications had been filed. However, indirect evidence could sometimes be obtained from related patents issued to the patentee or assignee. The U.S. law with regard to the publication of pending patent applications is expected to change in 1996, but final details are uncertain.

Under U.S. law, the inventor is defined as the owner of the patent unless the patent rights have been assigned to his or her employer, or some other individual or organization. Designations of assignment are typically filed with the U.S. Patent and Trademark Office (U.S. PTO) prior to the issuance of patents, and the name of the assignee is printed on the patent. In most countries outside of the United States, the patentee is the employer, rather than the employed inventor.

As the number of patent applications filed during the middle of the twentieth century grew, the time required to notify the public that an invention had been claimed in a pending patent application was seen as a serious inconvenience. Laws were introduced in some countries to inform the public about potential patents during their pendency. In some countries, the name of the applicant, the title of the patent application, and the serial number are published immediately after the filing of the patent application. The full patent specification is published or made "open for public inspection" (OPI) by most modern patent offices approximately 18 months after the original filing date of the application. The first publication of the patent application usually follows some initial examination of the application by patent examiners, but no judgment as to the patentability of the claimed invention is made then and the published specification will not ordinarily have been amended. Full examination of the patent application is omitted by some countries unless the validity of the claims is challenged by a third party. Most countries, however, proceed with the examination and publish the amended specification for a second time when the patent is eventually granted. The validity of the allowed patent claims can be challenged in formal opposition proceedings either before or after the formal grant of the patent. If the opposition proceedings result in modification of the patent claims, an amended granted patent may be published.

Once granted, patents are in force for a term prescribed by law. Patent terms are not renewable. Most countries have established a term of 20 years, measured from their national filing date, but patent laws enacted before the latter third of the twentieth century vary considerably in the length of the patent term. Japanese patents have a term of 20 years from filing, but until 1995, the term was subject to the limitation that it would expire no later than 15 years from the date of grant. Under U.S. law, patents based on applications made before June 8, 1995 have an effective term of 17 years from their grant date, regardless of the length of the interval between filing and issue. Patents filed on or later than June 8, 1995 have a term of 20 years from their original U.S. filing date. Those patents granted under the earlier law and still in force on June 8, 1995 have been given a revised expiration date of 20 years after their earliest U.S. filing if the regular 17-year term would cause them to expire earlier than that. The new U.S. law will also permit the filing of incomplete provisional applications, which do not require proposed claims, as domestic priorities. A regular application must be made within a year from the filing of a provisional application. The domestic priority period will not count in the life of an issued patent derived from a provisional application. Patents issued by the United Kingdom before the current 20-year term was established had a term of 16 years from filing. Terms of 16 and 17 years were once common, but have largely been supplanted by 20-year terms.

The expiration date of a patent is not normally printed on its face and must be calculated on the basis of the applicable national laws. Exceptions can occur in the United States when a term is foreshortened because of that patent's close relationship to a previously issued patent or, under the new law, when a term is extended because of delays in the course of patent prosecution. In addition, patents issued by most countries are kept in force by payment of periodic maintenance fees. Because some products cannot be marketed without the approval of governmental regulatory agencies, the owners of patents on drugs, medical devices, and agricultural chemicals have long complained that the effective term of their patents is less than the term of unregulated products. Some countries have provisions for the extension of the patent term for products approved for marketing under regulatory laws. Patent term extensions granted by the United States and Supplementary Protection Certificates granted by EPO member states are effective only for the approved product and not for other products that might be covered by the patent. To determine whether a particular patent is in force, it is necessary to obtain information about the current legal status of the patent from sources other than the patent specification itself. A compilation of national and international laws regarding patent expiration has been published by Derwent Information Ltd.

Patenting Procedures. Procedural pathways followed by patent applications filed in various countries and resulting in the publication of patent documents are shown in Figure 1. National filing of a patent application in the home country of the applicant is typically treated as priority filing under the Paris Convention for the Protection of Industrial Property. Patent applications may be filed directly in each country of interest to the applicant, or a single application under the Patent Cooperation Treaty may be filed to facilitate acquisition of patent rights in many countries. If the PCT is chosen, a preliminary examination is performed and the patent specification is published with a search report prior to the applicant's filing in each designated patent office, where national procedures are followed. Countries such as South Africa and Belgium follow the simplest procedures, examining the application for formal compliance with the patent laws and publish the specification as a granted patent without examining the claims for novelty and nonobviousness. In such countries, the validity of the patent is tested only when the patent is challenged in the courts by a third party. Long-standing U.S. practice withholds publication until all aspects of examination have been completed and the patent rights granted. New procedures for publication before grant have been promised, but details have not yet been established. No opposition by third parties is provided for in the United States, although on rare occasions a patent may be reissued to correct irregularities in patent prosecution or reexamined to permit reconsideration of prior art overlooked before the grant of the patent.

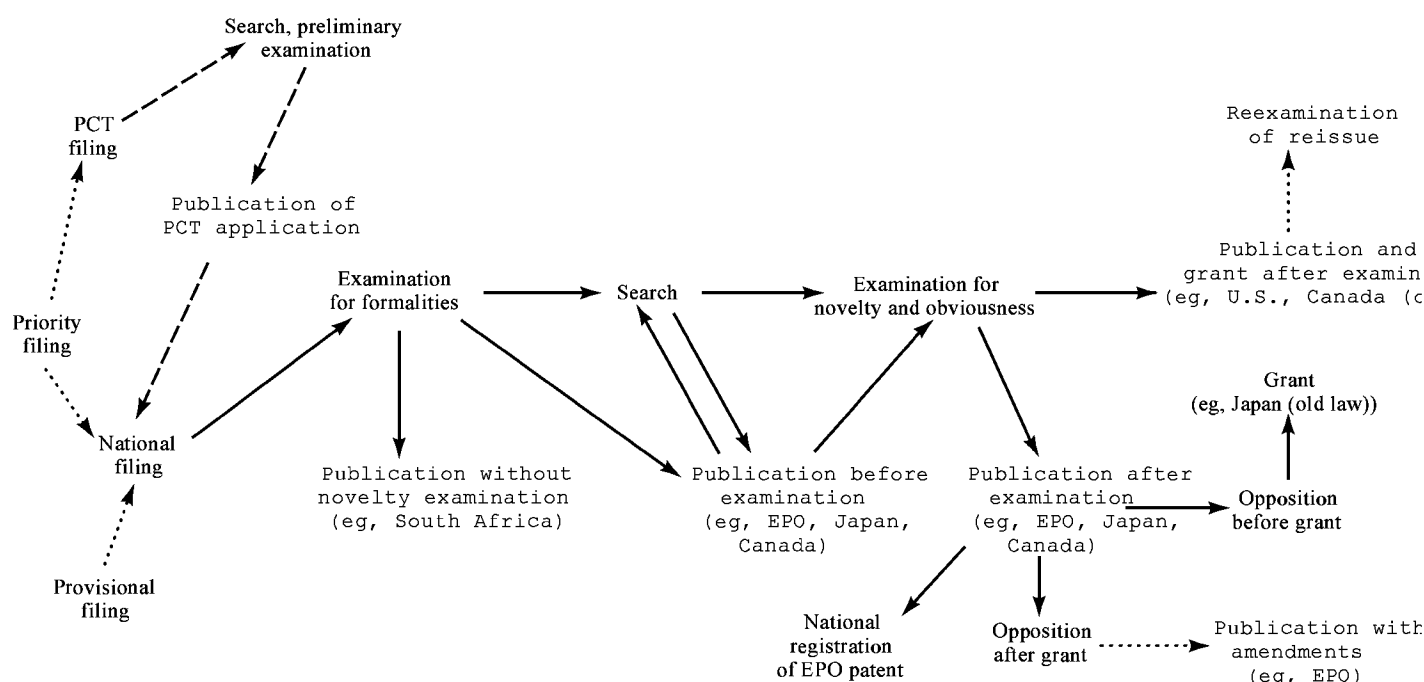


Fig. 1. Procedures for publication of patent documents. Publications are shown in boldface. Dotted lines indicate events that may take place before a national application has been filed or after a patent is granted. Dashed lines indicate events that take place only when PCT filing is chosen.

Many countries follow a more complex procedure. The patent specification is published before a substantive examination of the patent application. In Japan, publication takes place before a search of the prior art has been performed. In the European Patent Office (EPO), a search is made before the specification is published, and the search report is published at the same time or shortly afterward to assist the applicant in deciding whether to continue prosecution. If the claims are determined to be patentable by the patent office, the patent application is granted and opened to opposition by third parties. An amended patent will be issued if the scope of the patent is changed during opposition proceedings. In Japan prior to 1996, the opposition period took place before patent rights were granted. Beginning in 1996, examination will be followed directly by grant, as in the EPO, and the opposition period will take place after grant. Legislative and procedural changes take place from time to time; most countries establishing or revising patent laws are now adopting laws providing for terms of 20 years with publication 18 months after filing. For example, Canadian patent law paralleled that of the United States until 1989, when the former adopted a new patent law based on EPO procedures.

Patent Documents. The internal structure of patent documents has been standardized and the amount of bibliographic detail recorded in a patent document has increased. Early patent documents included rudimentary information about the patent's filing details. Most patent documents published during the 1990s begin with an informative cover page; an example is shown in Figure 2. The front page of a modern patent provides key information about the patent that aids the reader greatly in determining the patent's potential relevance. The cover page provides a title, gives the name of the patent owner, inventors, and other individuals involved in the issuance of the patent, and offers serial numbers and dates that identify the document and relate it to other patent documents covering the same invention. An abstract is provided by the patentee. Where appropriate, the abstract may include structural diagrams for chemical species important to the invention, and a representative drawing. National and International Patent Classification appropriate to the patent are shown, as is the list of classes searched by the examiner in determining patentability. Patent and other publications deemed by the examiner to be related to the invention are listed, and are referred to as examiner's citations. These bibliographic data have been standardized according to Internationally agreed Numbers for Identification of Data (INID) codes established by WIPO. The INID codes provide a means whereby the various data appearing on the first page of a patent and other similar documents can be identified without knowledge of the language used and the laws applied. They are used by most patent offices and have been applied to U.S. patents since August 4, 1970 (1).

United States Patent [19]
Himeno et al.

[11] **Patent Number:** **5,290,931**
[45] **Date of Patent:** **Mar. 1, 1994**
**[54] WATER-INSOLUBLE NAPHTHALIC ACID
IMIDE DYESTUFFS**

[75] Inventors: **Kiyoshi Himeno, Munakata; Toshio Hihara, Kitakyushu, all of Japan**
[73] Assignee: **Hoechst Mitsubishi Kasei Co., Ltd., Tokyo, Japan**
[21] Appl. No.: **940,789**
[22] Filed: **Sep. 4, 1992**

Related U.S. Application Data

[63] Continuation of Ser. No. 658,677, Feb. 21, 1991, abandoned.

[30] Foreign Application Priority Data

Feb. 21, 1990 [JP] Japan.....2-40302
Mar. 7, 1990 [JP] Japan.....2-56212

[51] Int. Cl.⁵**C07D 401/14; C07D 403/14; C07D 413/14; C07D 471/22**

[52] U.S. Cl.**544/112; 546/41; 544/113; 544/120; 544/122; 544/125; 544/182; 544/212; 544/218; 544/238; 544/298; 544/316; 544/319; 544/322; 544/326; 544/405**

[58] Field of Search**544/328, 125, 209, 212, 544/218, 238, 298, 322, 326, 319; 546/041**

[56] References Cited
U.S. PATENT DOCUMENTS

1,935,945 11/1933 Eckert et al.546/41
2,965,644 12/1960 Eckert et al.546/41
3,031,252 4/1962 Staeuble et al.544/209
3,183,231 5/1965 Buell544/209
3,518,266 6/1970 Häusermann et al.544/113
3,632,589 1/1972 Mueller et al.546/41
3,741,971 6/1973 Fuchs et al.546/47
3,748,330 7/1973 Fuchs et al.544/198
3,812,051 5/1974 Merkle et al.549/231
3,883,523 5/1975 Parton544/209
3,888,862 6/1975 Meininger et al.544/238
3,931,186 1/1976 Fuchs et al.546/41
4,051,134 9/1977 Friedrich et al.544/212

FOREIGN PATENT DOCUMENTS

1569761 6/1970 Fed. Rep. of Germany
1569737 9/1970 Fed. Rep. of Germany
2089618 1/1972 France
2144737 2/1973 France
250052 12/1985 Japan
1391324 4/1975 United Kingdom

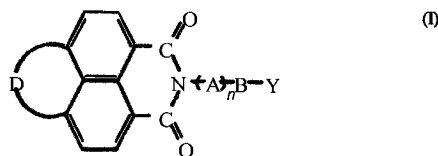
OTHER PUBLICATIONS

Chem. Abstr. 104: 208, 328 of JP-60-250,052 to Niwa et al. published Dec. 10, 1985.
Chem. Abstr. 78: 59, 785 of DE 2,122,975 to Hoechst published Nov. 1972.

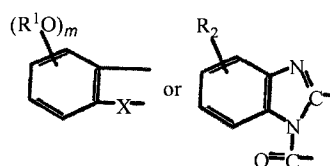
Primary Examiner—Nicholas S. Rizzo
Assistant Examiner—Matthew V. Grumbling
Attorney, Agent, or Firm—Obalon, Spivak, McClelland, Maier & Neustadt

[57] ABSTRACT

Water-insoluble naphthalic acid imide dyestuffs of the following formula:



wherein D is



wherein R¹ is C₁–C₈ alkyl which may be substituted, R² is hydrogen, halogen, lower alkyl, low alkoxy or cyano, X is oxygen or sulfur, and m is 0 or 1; A is phenylene, aralkylene, aralkylene or —C₂H₄OC₂H₄—; B is oxygen, sulfur or



wherein R³ is hydrogen or lower alkyl; Y is a 6-membered nitrogen-containing heterocyclic ring having at least one active halogen atom; and n is 0 or 1.

17 Claims, No Drawings

Fig. 2. Representative front page of a U.S. patent. The bracketed numbers are INID Codes. For example, [54] designates the title of an invention.

Because each country has its own patent laws, the precise meaning of the bibliographic data and the legal significance of the published patent document vary from country to country. The Patent Cooperation Treaty (PCT) provides a recommended code to distinguish the various types of documents and to simplify storage and retrieval of patent data (2), but the code is implemented differently by different countries. For example, in the United States an A-document in 1995 was a patent; in the Netherlands, an A-document was a published unexamined application. It is essential to understand each country's system to interpret the status of its patent documents.

The invention covered by the patent is defined in the patent claims, which appear at the end of patents by most countries and at the beginning of patents published by a few others. The majority of patents are known as utility patents. The claims of these patents may relate to new products, including new chemical compounds and compositions, to processes for making or using new or previously known products, and to machines for making or using such processes. Patent claims are examined by the national or international patenting authority to determine whether the claimed subject matter is patentable under the applicable statutes and by comparison with the prior art in the field of the invention. Where the claims are judged to cover more than one invention, the patent examiner may restrict the claims to a single inventive entity and authorize the filing of divisional patent applications. Where the inventor wishes to modify the invention, continuation-in-part applications or applications for patents of addition may be filed. What may be claimed in a patent differs from country to country, and has changed over time with amendments of national patent laws. Increasingly, patents are granted only when the claimed invention is novel, ie, has never before been patented, practiced, or described in a published document anywhere in the world; possesses an inventive step or is not obvious in light of the prior art; and has utility for a purpose acceptable under the law. Novelty and nonobviousness are determined by performing a search of the published patent and nonpatent literature. Many countries have limited patent rights to inventions considered socially useful or having industrial applicability. Methods of treating the human body, foods, pharmaceutical compositions, chemical compounds per se, living organisms, atomic weapons, computer programs, and scientific theories have all been held to be unpatentable in many countries at various times. Although the claims of granted patents are strictly limited by the various national laws, patent applications published before grant often claim subject matter that is not patentable; consequently, the claims of granted patents often differ significantly from the claims published in the unexamined application.

In addition to utility patents, some countries publish patent documents under different or less stringent standards for patentability and with shorter patent terms. For example, plant patents cover asexually reproduced plants. Design patents cover the decorative aspects of a product. Utility models and petty patents cover products with differences from the prior art that need not meet the nonobviousness standards set for utility patents.

The bulk of the patent specification is the disclosure, the text and illustrations that describe the claimed invention in detail and explain how the claimed invention differs from the prior art. Modern patent disclosures contain a summary of the claimed invention, a description of the background of the invention, a general description of the way in which the invention is made and used, specific examples, and, where applicable, drawings of the invention in general or specific embodiments. The technical information provided in a patent specification may be used without infringing the patent; only practicing the invention defined in the claims within the term and territory of the patent grant is forbidden. Because much of the information in patent specifications is never published in refereed journals or other nonpatent media, patent disclosures are an invaluable part of the technical literature.

Patent documents differ from journal literature in several ways. First of all, they are legal documents whose disclosures support one or more claims that define an area of property rights. The language in patent documents can therefore be quite convoluted "patentese" as the applicant strives to achieve the broadest possible scope of coverage. Examples provided in patents may never have happened. Based on the applicant's understanding of the technical

area involved, he or she may assume the probable outcome of experiments never actually run, and include in the patent specification such paper examples. Paper examples are generally written in the present tense. They lack hard data, and can provide grounds for attacking the patent should they prove to be inoperable. Finally, chemical patent disclosures and claims can be written in terms of generic structures, or the so-called Markush structures, in which one or more portions of a chemical entity can vary, including functional groups, numbers of substituents, and points of attachment. Markush structures are used as one method of obtaining the broadest possible claims in a patent, and are named after an early inventor who succeeded in obtaining claims on a process for making such variable products. Markush structures can be simple, describing just a handful of chemical compounds, or highly complex, encompassing thousands, millions, even infinite numbers of compounds; a typical example is shown in Figure 3. The effective indexing and searching of Markush structures provides a significant challenge to those concerned with chemical patents (3).

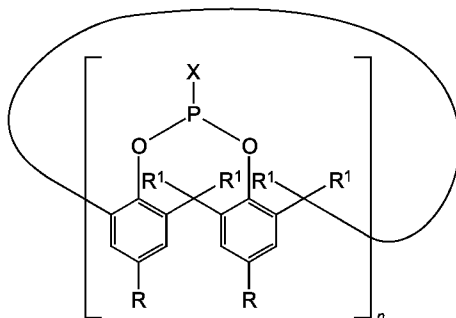


Fig. 3. Typical Markush structure where R is H, C₁₋₂₀ alkyl, C₁₋₂₀ alkoxy, sulfonate, or carboxylate; R¹ is H or C₁₋₂₀ alkyl; and X is H, C₁₋₂₀ alkyl, phenyl (Ph) or OPh, the last two of which may be substituted by up to three groups selected from C₁₋₂₀ alkyl, C₁₋₂₀ alkoxy, sulfonate, carboxylate, C₁₋₂₀ alkylthio-, and or C₂₋₂₀ dialkylamino- (3).

Patent Families. Patent specifications are published as individual documents in the language of the originating country, but many inventions are claimed in patents issued by more than one country. These patents form a family of equivalent patent documents, which usually disclose the same information but may differ somewhat in the scope of their claims. When filing in more than one country, an applicant establishing priority under the Paris Convention is generally required to submit a copy of the original application to each national or regional patent office selected, sometimes with a translation. A simple patent family is based on a single priority application, in which each family member discloses the same information and cites the same priority application number. When the technical content of the patent rather than its legal scope is of interest, any member of the family can be substituted for another, thus often obviating the need for translation.

When an applicant misses the deadline for convention-filing or files applications in countries that are not members of the Paris Convention, the priority application number will not be present in all patent applications having equivalent disclosures and claims. Nonconvention equivalents, which can only be recognized by comparing the contents of the applications, form what WIPO defines as an artificial, intellectual, technical, or nonconvention family. When the applicant has refiled the original national patent application within the priority year, more than one priority application may be claimed in the corresponding foreign applications. Applicants are also permitted to combine the disclosures of two or more patent applications within the priority year and file foreign counterparts incorporating information from each. A complex patent family contains patent documents having at least one common originating application. An extended patent family contains all of the patent documents having at least one priority application in common with any other member of the extended family. Extended patent families sometimes include patent documents that differ radically from other family members. Members of a national patent family, one that includes divisional, continuation, continuation-in-part, and addition patents issued in a single country, may also differ significantly in content. Patent databases usually provide technical information from a representative member of a patent family, which defines the patent family according to one of the definitions above or devises a hybrid definition unique to the database.

Patent Searches. Because valid patent claims can only be issued on an invention that is novel and innovative in light of prior art, it is necessary to search the prior art for previous references either to the composition of matter, process, or machine defined in the claims of a patent application, or to any similar composition, process, or apparatus that would render the claimed invention obvious to a person skilled in the field of the invention. Inventions that have been described in a publication or embodied in a product are said to have been anticipated in the prior art and are not patentable. Patentability searches are performed by examiners employed by the national and regional patent offices and are an important step in the examination of patent applications. Patentability searches should also be performed by the representatives of inventors prior to the filing of a patent application so that the claims will not overlap with any publication in the prior art. These searches may encompass the full scope of the published literature, including patents, technical journals, gray literature, and even catalogs. Individuals or organizations who are making plans to introduce a new product or process must conduct infringement searches to ensure that they will not infringe patents that belong to others. Infringement searches need only consider patents in force and pending applications that may result in patents in countries where manufacturing or marketing are contemplated. After a patent application has been published and or a patent has been granted, organizations that wish to practice the invention may also conduct validity searches to be used as ammunition for opposition proceedings or invalidity lawsuits. Validity searches, like patentability searches, should include all forms of published literature, but are limited to publications with effective dates earlier than the filing date of the patent application being challenged.

Searches of scientific and technical literature are performed using any of the information retrieval tools suitable for searches done for other purposes (see INFORMATION RETRIEVAL). Patent offices have devised special classification systems to facilitate searches among the individual patent documents in their collections. These patent classification systems were designed to subdivide patents into groups covering similar inventions that could be reviewed by examiners when related inventions were claimed in later applications. All of the existing fields of science and technology were defined and provided with a class code and subdivisions of the fields were given narrower classification designations. Patents belonging to each subclass were stacked together in drawers or on shelves similar to the stacks of boxes in a shoestore, and examiners or members of the public could extract a stack of patents and search for information in the subfield of interest by flipping through paper copies of the patent documents. As new fields of science and technology have developed, each patent classification system has been revised so that the emerging technologies can be searched. Patents are assigned classification codes by the examining office and the relevant primary classification and any cross-reference classifications are printed on the first page of the patent, eg, INID codes [51] and [52] in Figure 2. Although patent classifications originated as tools for manual searches, they can be searched through printed or electronic indexes as well.

National patent offices created patent classification systems for internal use without correlating their guidelines for subdividing technologies or the symbols used to identify classifications with those of other countries. The assignment of national classification codes to patents of the issuing country facilitates manual searching for inventions claimed in national patents, but is not helpful for prior art searches that must include patents issued by patent offices that use different classification systems. However, the internationalization of commerce has led to the internationalization of patent classifications. The first edition of the International Patent Classification (IPC) system was introduced in 1968; it has been revised by WIPO every five years and is in its sixth edition in 1995 (4). The IPC has been adopted by most of the patent-issuing countries of the world. Even countries such as the United States, which continues to use a national classification system to organize their patent search files, print a corresponding IPC classification on the patent documents. Although the IPC is used by most countries, these countries do not all follow the same guidelines for applying the codes, nor do they all use the finest divisions of the classification system. Differences in the scope of the patent claims, on which the IPC classification is based, as well as differences in the classifying examiner's interpretation of the novel features of the invention, also contribute to differences in patent classification among countries. It is not unusual for patents having identical claims to be classified differently in each country where the patent application was filed.

IPC codes, which have the format *ANNA NNN/NN*, where *A* stands for a letter and *N* a numeral, represent a hierarchical system (4). The first four characters designate the section, class, and subclass of the class code, and each successive character narrows the definition of the invention. Each subclass is further divided into groups, defined by one to three numerals and followed by two- to three-digit subgroup designations. The hierarchical relationships within the groups are determined by the relationships published in the IPC manual (Fig. 4). Some countries index patents only to the four-character subclass level, whereas others use the full IPC. Although a main IPC code is always provided, some countries assign supplementary and or additional IPC codes that designate additional aspects of the claimed invention. In searching, it is usually necessary to truncate IPC codes unless the level of specificity used by the country of interest is known.

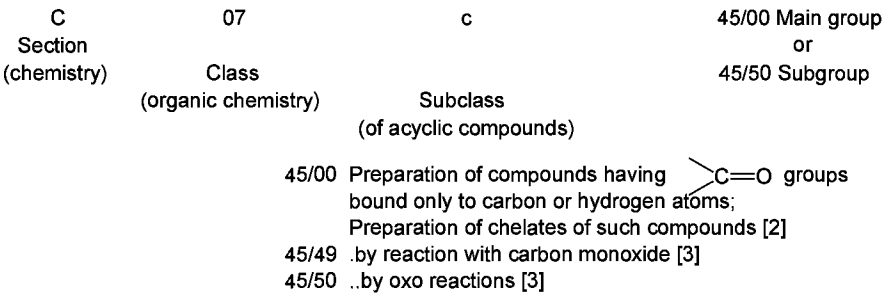


Fig. 4. Example of international patent classification (structured, hierarchical), where numbers in square brackets identify edition of IPC in which class was first used. In C07c 45 50, the first four characters indicate section C (chemistry), Class 07 (organic chemistry), and subclass c (acyclic compounds); the number 45 00 indicates the preparation of compounds having carbonyl groups bound only to carbon or hydrogen atoms by any method; and 45 50 indicates such preparation by oxo reactions.

A hybrid system of classification has been introduced for some specific indexing of patents; countries that use the hybrid system append indexing codes in the format *NNN:NN* to IPC subclass codes for certain technologies. Some patent offices and databases, but not all, identify the edition of the IPC used in classifying a patent. Because variations in patent classification practices can cause difficulties in searching multinational patent files, the European Patent Office has reclassified all of the patents in its search documentation files according to its own standards, using a modified version of the IPC. The European Patent Classification (ECLA) scheme is based on the hierarchical structure of the IPC, but has additional subdivisions to allow more focused searching by European Patent Office staff (5). Monthly updates are made to the classification scheme to adapt the classification to technological development.

The most common national patent classification codes encountered outside the public search rooms of national patent offices are U.S. classes, which are indexed in many patent databases that include U.S. patents. They are formatted as a one- to three-digit numerical class code, followed by a slash or hyphen and a subclass code consisting of from one to three numbers, which are occasionally followed by a letter or by a decimal point and additional numerals. These codes are also arranged hierarchically according to the scheme published in the U.S. *Manual of Classification* (6,7). Unlike IPC codes, U.S. patent classification codes do not contain clues to their technological relationships. Patents are given a single original classification and usually one or more cross-reference classifications. The U.S. patent classification system is under constant revision and, as the purpose of the system is to provide a useful arrangement of the patents on the shelves of search rooms, file copies of the patents are moved to their new places at U.S. PTO search rooms, and indexes are revised when new classifications are assigned to existing patents. The printed patent specification is not changed when a patent is reclassified, but the U.S. Patent and Trademark Office records the current classification codes of reclassified patents and makes the data available to users of the public search room and to database producers. The U.S. system designation corresponding to the IPC C07c 45 50 would be 568-451 (Fig. 5).

CLASS 568	ORGANIC COMPOUNDS
568-300	.OXYGEN CONTAINING
420	..Aldehydes
448	...Acyclic
449	Processes
450	Isomerization
451	Hydroformylation by reacting ethylenically unsaturated compound, carbon monoxide, and gaseous hydrogen
452	Dimer produced
453	Plural stages each having hydroformylation
454	Group 15 (VA) element (N, P, As, Sb, or Bi) containing material utilized (eg, arsenic containing ligand utilized, etc)

Fig. 5. Example of U.S. patent classification (unstructured, hierarchical). This is one of a series of classes considered as integral parts of Class 260, following the schedule hierarchy retaining all pertinent definitions and class lines of Class 260.

Secondary Sources of Patent Information

Patent systems were conceived as a means for promoting technical progress by encouraging the dissemination of information on technological developments. Information dissemination is therefore essential for the patenting process. Patent offices have traditionally announced the issuance of new patents in bulletins and gazettes. Other organizations, notably scientific and technical societies and for-profit publishers, have produced value-added patent information services. These secondary sources of patent information serve multiple purposes, among which are current awareness alerting, document delivery, and retrospective searching. Traditionally, such products have appeared as printed publications, but increasingly they have found second use in electronic form in on-line databases, and in the 1990s there has been rapid growth of optical storage of information, especially as Compact Disk-Read Only Memory (CD-ROM) products. Patent documentation is a field in considerable ferment, with rapid introduction of new products and capabilities.

Printed Patent Office Gazettes. The issuance of patents is announced by patent offices in publications typically known as gazettes and bulletins, which are published most commonly at the time of the patent's publication, but there are exceptions. Advance information is published in a patent gazette by some countries prior to the publication of patent documents, typically as a notification of filing details. However, some patent gazettes do not appear until well after the effective publication date of the patents they announce.

The amount of information included in patent gazettes varies. Typically, they include bibliographic details on published patent applications and granted patents, including patent number, title, inventor, patentee, patent classification, application number and date, and priority application details if relevant. Some gazettes also provide the front page abstract of the patent and a representative drawing. Examples are the *PCT Gazette*, the *Bulletin Officiel de la Propriété Industrielle* of France, and the *Patent Journal*, or *Patentjoernaal*, of South Africa. The *Official Gazette* of the U.S. Patent and Trademark Office

includes one or more representative claims. In addition to announcements of new patents and applications, the various gazettes typically include listings of patents that have been rejected, challenged, or disclaimed, patents that have been allowed to lapse, and in some instances even listings of new applications that have been made but that will not be published for some time, if ever. Gazettes often include indexes to the information they contain; the amount of indexing available varies from country to country.

Information from Other Sources. Some of the abstracting and indexing services produced by scientific and technical societies have traditionally included patent information, especially in the field of chemistry. For instance, *Chemical Abstracts* (CA), produced by the American Chemical Society since 1907, has always covered patents, as did the discontinued *Chemisches Zentralblatt* and *British Chemical Abstracts*. On the other hand, some notable information services have not included patent coverage. One example, despite the fact that many patents are based on some aspects of engineering, is the *Engineering Index. Science Abstracts*, covering physics, electricity, and electronics, is another example, which has not covered patents since 1976. *RAPRA Abstracts*, focusing on rubber and plastics, covered patents briefly from 1978–1980, dropped them for more than a decade, and finally resumed the coverage in 1994. However, even where patents are covered, the focus may not be ideal for those concerned with the legal aspects of patents. Thus, CA in its patent coverage documents the new chemistry involved, but shies away from the legal aspects of patents. For these and other reasons, others have stepped in to develop a variety of patent information services, eg, Derwent Information Ltd. of London.

Derwent had its start in the 1950s, when it began publishing abstracts of patents from selected countries: first the United Kingdom, then Belgium, followed by others that included Japan, the Netherlands, and the former USSR. Derwent's country abstract booklets were followed by collections of abstracts covering multiple countries in selected technical areas. These products served an alerting purpose, but had no capability for retrospective searching. During the 1960s, Derwent began a series of more complex information services, providing both alerting and retrospective retrieval capabilities in the fields of pharmaceuticals, agricultural chemicals, and polymers. By 1970, Derwent's coverage was extended to all aspects of chemistry from 12 countries; by 1974, it had begun coverage of nonchemical patents. Over the course of the ensuing years, the Derwent organization has broadened its country coverage and improved its capabilities for information retrieval in many ways. It continues to work on new and improved products and systems, and is the single most important organization involved in patent documentation.

Other organizations have assumed important positions in the field of patent documentation. IFI Plenum Data Corp. (formerly Information for Industry) began in 1955 to index U.S. chemical patents by the Uniterm Index system. Uniterm indexing was eventually extended back to 1950. The acquisition in 1971 of Du Pont's in-house indexing system and staff resulted in a more powerful system, the Comprehensive Data Base (CDB), which now covers U.S. chemical patents from mid-1964 to date.

Another important resource for bibliographic information on patents is the European Patent Information and Documentation Systems (EPIDOS). EPIDOS began in 1973 as the International Patent Documentation Center (INPADOC), the joint creation of the Austrian government and WIPO. Subsequently acquired by the European Patent Office, EPIDOS continues to produce the INPADOC database, which at present covers patents from about 60 authorities, and is the most complete source of bibliographic information on patents, including patent family and legal status information.

There are other organizations providing patent information. L'Institut National de la Propri  t   Industrielle (INPI), the French Patent Office, is the producer of several important databases and, together with Derwent and the Questel databank, has supported the development of the Markush DARC system used in Derwent's WPIM and INPI's PHARMSEARCH databases. The American Petroleum Institute's (API) Central Abstracting and Information Service has since 1964 produced APIPAT, a database covering patents on petroleum refining, petrochemicals, and related technology. API also produces APILIT and APIBIZ, complementary databases covering published technical literature and business information.

Other specialized patent information products, as well as general information products that include patent and other information, are produced by a variety of organizations, most notably in the area of pharmaceuticals. Unlike most other products, new drugs must undergo extensive regulatory evaluation before they are marketed. Patent terms and terms of market exclusivity for approved drugs are of paramount importance to manufacturers and their potential competitors. Lists of patents relied on for the protection of drugs marketed in the United States are published by the U.S. Food and Drug Administration (FDA) in *Approved New Drugs*, also known as the Orange Book, and the FOI Services Inc. in the series *Drugs Under Patent*. Patents on new veterinary drugs are listed in a corresponding Green Book published by the FDA. On-line searches for patents covering drugs may be performed in the IMSWorld Drug Patents international database. This database contains patent records from many countries on about 1000 marketed drugs and, like the printed lists, is searched by the name of the product. Although these product-specific sources are rather limited in their scope, they provide precisely the patent information most sought by pharmaceutical company executives.

Advances in Patent Documentation

The last half of the twentieth century has seen a strengthening of patent coverage by some traditional abstracting and indexing services whose patent coverage extends back for many years, as well as the establishment of an increasing number of specialty services for the documentation and manipulation of patent information. Advances have involved traditional printed products as well as various electronic forms of information. Computerized databases have become increasingly important to users of patent information, and new and modified information tools continue to appear and develop. Principal patent databases available through on-line databanks are listed in Table 2.

Table 2. Patent Databases

Database	Systems	Producer	Coverage
World Patents Index	DIALOG, ORBIT, Questel, STN	Derwent Information Ltd.	international; limited bibliographic data; patent families; comprehensive English language abstracts of basic and some equivalent granted patents; polymer and chemical structure indexing for subscribers; drawings
INPADOC	INPADOC, PATOLIS, DIALOG, ORBIT, STN	International Patent Documentation Center	international; bibliographic data; patent families; legal status data for 16 countries
EDOC	Questel	Institut National de la Propri��t�� Industrielle (INPI)	international; minimal bibliographic data; patent families; European Patent Classification codes
U.S. Patents Fulltext (PATFULL)	DIALOG	Knight Ridder Information Services; IFI Plenum Data Co.	U.S.; full bibliographic data; full text; some controlled indexing from CLAIMS
USPATFULL	STN	Chemical Abstracts Service	U.S.; full bibliographic data; full text; <i>Chemical Abstracts</i> indexing from U.S. patent or equivalent
LEXPAT	LEXIS–NEXIS	LEXIS–NEXIS	U.S.; full bibliographic data; full text; post-grant status changes
CLAIMS	DIALOG, ORBIT	IFI Plenum Data Co.	U.S.; full bibliographic data; abstract and full claim text; chemical structure, general concept and patentee name coding
USPatents	ORBIT	Derwent, Inc.	U.S.; full bibliographic data; abstract and full claim text
European Patents	DIALOG, STN	European Patent Office	EPO; full bibliographic data; abstract and first claim text in English, French, and German; full text of published application and granted patent in original language planned; prosecution and status data
European Patent Register	EPOIS	European Patent Office	EPO; full bibliographic data; abstract and first claim text in English, French, and German; prosecution and status data

EPAT	Questel, ORBIT	Institut National de la Propri��t�� Industrielle	EPO; full bibliographic data; abstract and first claim text; status data
PATOSEP	STN	Wila Verlag; Bertelsmann Information Service	EPO; full bibliographic data; abstract and first claim text in German and English; status data
PCTPAT	Questel	Institut National de la Propri��t�� Industrielle; World Intellectual Property Organization	PCT; full bibliographic data; abstract and first claim text; status data
PATOSWO	STN	Wila Verlag; Bertelsmann Information Service	PCT; full bibliographic data; abstract and first claim text; status data
FPAT	Questel	Institut National de la Propri��t�� Industrielle	France; full bibliographic data; abstract and first claim text; status data
ITALPAT	Questel	JUSTINFO Ltd.	Italy; minimal bibliographic data for applications; Italian title
PATOSDE	STN	Wila Verlag; Bertelsmann Information Service	Germany; full bibliographic data; abstract and first claim text; status data
PATDPA	STN	Deutsches Patentamt	Germany; full bibliographic data; abstract and first claim text; status data; exemplary drawing
PATDD	STN	Deutsches Patentamt	GDR; bibliographic data and abstract
Chinese Patent Abstracts	DIALOG, ORBIT	European Patent Office	China; bibliographic data and English language abstract
PATOLIS	PATOLIS	Japan Patent Information Organization	Japan; full bibliographic data; abstract and first claim text in Japanese; status data; drawings
JAPIO	DIALOG, ORBIT, Questel, STN	Japan Patent Information Organization	Japan; bibliographic data and English language abstract
Patent Citation Index	DIALOG, Questel��ORBIT, STN	Derwent Information Ltd.	international; examiners' and inventors' citations of earlier references and later citing patents; Derwent title, bibliographic, and family information
CLAIMS-Citation	DIALOG	Search Check, Inc.; IFI Plenum Data Corp.	U.S.; patent numbers; examiners' citations of earlier patents; later citing U.S. patents
APIPAT	DIALOG, STN	American Petroleum Institute	international, petroleum, petrochemical; limited bibliographic data; comprehensive English language abstracts of basic patents; concepts and chemical structure coding for subscribers
World Patents Index APIPAT	ORBIT	Derwent Information Ltd.; American Petroleum Institute	international; limited bibliographic data; patent families; comprehensive English language abstracts of basic patents; all World Patents Index and APIPAT indexing for subscribers
CA File, CAPlus, CA Previews	STN	Chemical Abstracts Service	international, chemistry; limited bibliographic data; comprehensive English language abstracts; deep indexing of chemical concepts; structure-searchable compound registry; chemical structure drawings
MARPAT, MARPAT Previews	STN	Chemical Abstracts Service	international, chemistry; structure-searchable Markush formulas in addition to other data searchable in the CA File, CAPlus, or CA Previews
CA Search	Data-Star, DIALOG, ORBIT, Questel, etc	Chemical Abstracts Service	international, chemistry; limited bibliographic data; deep indexing of chemical concepts; compound registry searchable by structure on Questel and by name and molecular features on other systems
Biotechnology Abstracts	Data-Star, DIALOG, ORBIT, STN	Derwent Information Ltd.	international, biotechnology; limited bibliographic data; comprehensive English language abstracts of basic patents
GENESEQ	IntelliGenetics, STN	Derwent Information Ltd.	international, biotechnology; limited bibliographic data; polypeptide and nucleic acid sequences and related indexing
PHARM-SEARCH	Questel	Institut National de la Propri��t�� Industrielle	international, pharmaceutical; bibliographic data; English language abstract and pharmaceutical indexing; chemical structure drawings
Current Patents	Data-Star, ORBIT	Current Drugs Ltd.	international, pharmaceutical; bibliographic data; English language abstract focused on pharmaceutical utility
Drug Patents Interna-tional	Data-Star, DIALOG, ORBIT, STN	IMSWorld Publications Ltd.	international, marketed pharmaceuticals; bibliographic data; information relating to drugs covered by patents

Countries covered by patent databases vary from one database to another; some databases provide complete information about patents published by a single patent-issuing authority, others attempt to catalog the world's entire patent output. Multinational patent databases have historically provided good coverage of heavily industrialized countries and lesser coverage of less industrialized countries. The breadth of coverage and depth of indexing for these databases have depended on the availability of original source materials such as published patent applications, published patent office gazettes, and computerized records provided by the patent offices. Coverage has also been limited by the expense of translating information from patents in subject areas for which the market is small. Country coverage has tended to expand and contract in response to consumer demand and as the availability of documentation from patent offices changes for political and procedural reasons. A summary of the countries covered by major computerized patent databases and of the country codes identifying the countries is given in Table 3. Coverage for individual countries may vary in thoroughness with respect to subject matter, type of document, and time span. Users' manuals from such database producers as Derwent and EPIDOS include detailed lists of country coverage ranges.

Table 3. Country Coverage of Patent Databases

Country ^a	Treaty ^b	ISO code ^c	Databases and years of coverage					
			WPI	INPADOC	EDOC ^d	APIPAT ^e	CA Search	Other databases
Albania		AL						
Argentina		AR	1975��	1973��				
ARIPO ^f		AP		1984��				
Armenia		AM						
Australia		AU	1963��1960	1973��	1973��		1967��	

Austria	EPO	AT [OE]	1982– 1975–	1969–	1971–	1967–	
Azerbaijan		AZ					
Barbados		BB [BD]					
Belarus		BY					
Belgium	EPO	BE	1963–	1964–	1964–	1964–	1967– CLAIMS: 1950–1979
Benin	OAPI	BJ [DA]					
Brazil		BR	1976–	1973–			1976–
Bulgaria		BG		1973–			
Burkina Faso	OAPI	BF [HV, UV]					
Cameroon	OAPI	CM [KA]					
Canada		CA	1963–	1970–	1970–	1964–	1967–
Central African Republic	OAPI	CF [ZF]					
Chad	OAPI	TD [TS]					
China		CN	1985–	1985–			CHINAPATS: 1985–
Congo	OAPI	CG [GF]					
Cuba		CU		1974–			
Cyprus		CY		1975–			
Czech Republic		CZ	1993–	1993–			
Czechoslovakia		CS	1975–1992	1973–1992			1967–
Denmark	EPO	DK	1974–	1968–			1967–
Egypt		EG [ET]		1976–			
Estonia		EE					
Europe		EP	1978–	1978–	1978–	1978–	1979– EPAT: 1978– EPOIS: 1978– PATOSEP: 1978– European patents: 1978– PHARMSEARCH: 1985–
Finland		FI [SF]	1974–	1968–			1967–
France	EPO	FR	1963–	1968–	1968–	1964–	1967– FPAT: 1969– CLAIMS: 1950–1979 PHARMSEARCH: 1961–1973; 1985–
Gabon	OAPI	GA					
Georgia		GE					
GDR		DD [DL] (EG)	1963–	1973–		1983–	1967– PATDD: 1982–
FRG	EPO	DE [DT] (DS, GE)	1963–	1967–	1968–	1964–	1967– PATDPA: 1968– PATOSDE: 1968– CLAIMS: 1950–1979 PHARMSEARCH: 1992–
Greece	EPO	GR		1977–			
Guinea	OAPI	GW					
Hong Kong		HK		1976–			
Hungary		HU	1975–	1973–			1967–
Iceland		IS					
India		IN		1975–			1967–
Ireland	EPO	IE [EI]	1963–1969 1995–	1973–			
Israel		IL	1975–	1968–			1967–
Italy	EPO	IT	1966–1969 1977–	1973–			1967–1974 ITALPAT: 1983–
Ivory Coast	OAPI	CI					
Japan		JP (J#) [JA]	1963–	1973–	1973–	1968–	1967– JAPIO: 1976– PATOLIS: 1955–
Kazakstan		KZ					
Kenya	ARIPO	KE		1975–			
Korea, North		KP [KN]					
Korea, South		KR [KS]	1986–	1978–			
Kyrgyzstan		KG					
Latvia		LV					
Lesotho		LS					
Liberia	ARIPO	LR					
Liechtenstein	EPO	LI					
Lithuania		LT					
Luxembourg	EPO	LU	1984–	1960–	1961–		
Macedonia		MK					
(former Yugoslav Republic)							
Madagascar		MG [MD]					
Malawi	ARIPO	MW		1973–			
Mali	OAPI	ML [MJ]					
Malaysia		MY		1953–			
Malta		MT		1967–			
Mauritania	OAPI	MR [MT]					
Mexico		MX	1995–	1981–			
Moldova		MD					

Monaco	EPO	MC		1975–					
Mongolia		MN [MO]		1972–					
Netherlands	EPO	NL	1963–	1964–	1965–	1964–	1967–	CLAIMS: 1950–1979	
New Zealand		NZ	1992–	1978–					
Niger	OAPI	NE							
Norway		NO	1974–	1968–			1967–		
OAPI		OA				1964–			
PCT		WO (WP)	1978–	1978–	1978–	1978–	1979–	PCTPAT: 1978– PATOSWO: 1983 PHARMSEARCH: 1992–	
Philippines		PH [RP]	1994–	1975–					
Poland		PL [PO]		1973–			1967–		
Portugal	EPO	PT	1974–	1976–					
Romania		RO [RU]	1975–	1973–			1967–		
Russia		RU	1993–	1993–					
Senegal	OAPI	SN							
Singapore		SG	1995–	1983–					
Slovakia		SK	1993–						
Slovenia		SI		1994–					
South Africa		ZA (SA)	1963–	1971–		1964–	1968–		
Spain	EPO	ES	1983–	1968–			1967–		
Sri Lanka		LK [CL]							
Sudan	ARIPO	SD							
Swaziland	ARIPO	SZ							
Sweden	EPO	SE [SW]	1974–	1968–	1973–		1967–		
Switzerland	EPO	CH (SW)	1963–	1969–	1968–		1967–		
Taiwan		TW	1993–						
Tajikistan		TJ							
Trinidad and Tobago		TT							
Togo	OAPI	TG [TO]							
Turkey		TR		1973–					
Turkmenistan		TM							
Uganda	ARIPO	UG							
Ukraine		UA							
United Kingdom	EPO	GB (BR)	1963–	1969–	1968–	1964–	1967–	CLAIMS: 1950–1979 PHARMSEARCH: 1992–	
United States		US	1963–	1968–	1969–	1964–	1967–	CLAIMS: 1950– USPatents: 1970– LEXPAT: 1975– PATFULL: 1975– USPATFULL: 1975– PHARMSEARCH: 1985–	
USSR		SU (RU)	1963–	1972–		1983–	1967–		
Uzbekistan		UZ							
Vietnam		VN		1984–					
Yugoslavia		YU		1973–					
Zambia	ARIPO	ZM [ZB]		1968–					
Zimbabwe	ARIPO	ZW [RH]		1980–					

^a All countries listed are included in Patent Cooperation Treaty (PCT) except for Argentina, Cuba, Cypress, Egypt, the former GDR, Hong Kong, India, Israel, Malaysia, Malta, Philippines, South Africa, Taiwan, the former Yugoslavia, Zambia, and Zimbabwe.

^b Countries designated as members of the European Patent Organization (EPO), the African Intellectual Property Organization (OAPI), or PCT may be found as designated states in the records of EP, OA, and WO documents, whether or not those countries are indexed separately in a database.

^c Country codes in brackets are obsolete ICIREPAT codes. Country codes in parentheses were used by Derwent Publications Ltd. prior to the adoption of International Standards Organization (ISO) country codes. The obsolete codes remain on printed records, but have been replaced by current codes in on-line DATABASES. For Japanese patents, Derwent replaced the second character of the country code with the first digit of the publication year prior to 1992; this variant of the company code is still found in on-line databases and printed publications.

^d Countries not covered systematically include Brazil, the former Czechoslovakia, Denmark, Finland, the former GDR, Hungary, Israel, Italy, Norway, Portugal, Romania, South Africa, Spain, and the former USSR. In addition, patents from some of the systematically covered countries may have been indexed in earlier years.

^e Countries not covered systematically include Argentina, Australia, Austria, Brazil, the former Czechoslovakia, Denmark, Finland, Hungary, India, Israel, Italy, Norway, Poland, Portugal, Romania, Spain, Sweden, and Switzerland. In addition, patents from some of the systematically covered countries may have been indexed in earlier years.

^f Industrial Property Organization for Africa; also includes Botswana (BW), Gambia (GM), and Ghana (GH).

Derwent Information Ltd. Derwent Information Ltd., previously known as Derwent Publications Ltd., changed its name in the 1990s to reflect more accurately the fact that its products go far beyond traditional publications. Derwent provides a wide spectrum of information products and services, many of them relating to patents. Derwent also produces important databases (qv) of nonpatent information from the pharmaceutical and agricultural chemical literature. These products and services encompass alerting tools for current awareness, systems for retrospective search and retrieval, and means for document delivery and archiving.

Derwent began as a publisher of simple abstract booklets covering first individual countries, then multiple countries in specific areas of technology, and gradually moved into covering the full range of chemistry (in depth that differs from one subfield to another) and to nonchemical patents. The basic framework of the *Chemical Patents Index* (CPI) was established in 1970, and the overall *World Patents Index* (WPI), encompassing CPI as well as nonchemical patents, was established in 1974. The CPI is divided into 12 sections by technology, as shown in Table 4. The organization of the CPI is essentially unchanged in 1995, although it has undergone many refinements and its country coverage has been broadened considerably. About 40 patenting authorities

are covered in 1995, including European and PCT patent publications, and multiple stages of publication are covered for an increasing number of countries. The trend to broadened coverage is likely to continue. Limitations were placed on the coverage of Japanese patents during the 1970s in response to the sharp rise in numbers of Japanese documents after Japan had adopted the practice of universal publication of patent applications. Chemical coverage was not complete, electrical patents were covered by title only, and general and mechanical patents were excluded. However, Japanese coverage has gradually broadened since then, and extension to all Japanese patents is scheduled for the end of 1995. Not all countries are covered in equal depth. Thus, title-only coverage is provided for Italy and the Czech Republic, and abstracts for some other countries are briefer than the Derwent norm. Treatment of general and mechanical patents in the WPI is similar to that of the 1970s except for the addition of more countries; however, coverage of electrical patents in the *Electrical Patents Index* (EPI) segment of the database has been upgraded considerably with the establishment of a number of subdivided bulletin groupings and the development of an extensive system of manual codes, which have been upgraded on several occasions.

Table 4. Scope of Derwent CPI and WPI

Section	Subject content	Number of basic patent references, $\times 10^3$		
		1973	1983	1993
A	PLASDOC: polymers	28.3	39.7	63.1
B	FARMDOC: pharmaceuticals	6.8	10.1	18.9
C	AGDOC: agricultural chemicals	4.0	5.1	6.2
D	food, biotechnology, detergents, cosmetics, etc	8.3	15.5	29.4
E	CHEMDOC: general chemicals	15.8	19.6	27.8
F	textiles, paper, cellulose	11.6	10.1	15.1
G	printing, coating, photographic chemistry	5.4	10.0	21.2
H	petroleum	4.3	8.1	9.0
J	chemical engineering	6.6	14.4	19.6
K	nuclear, explosives, protection	2.2	3.4	4.3
L	glass, refractories, ceramics, electrochemistry	9.0	26.6	47.8
M	metallurgy	13.1	25.2	32.3
Total CPI		81.0	128.0	183.5
Total WPI		81.0	302.4	413.5

The Derwent patent database is based on records covering a family of equivalent patents. New patent publications are checked against the existing database to see if they are equivalent to previously published references. This is done by comparing priority application details or, if priority is not claimed, by comparing inventor or patentee and technical content with known references. Those publications determined to be new to the system are considered to be basic patents, and are assigned to one or more sections of the Derwent system according to their technical content. Chemical patents may appear in as many sections of the CPI as needed, although a limit of four sections per patent was applied in the past.

Table 4 shows the size of CPI and of its 12 sections. The annual input has more than doubled over the span of 20 years, and shifts in activity are evident. Thus Section L accounted for just over 11% of CPI references in 1973 but over 26% in 1993; Section D grew from 10 to 16%, and Section G from 6.6 to 11.5%. The largest relative declines occurred in Sections E and F, while Section A, the largest of all chemical sections, remained relatively constant at about 33% of the CPI. The sectional breakdown of CPI has been important for the marketing of the service. The cost of purchasing coverage for all technologies is substantial, and few organizations have interests sufficiently broad to buy the complete service. By packaging the product in segments, Derwent has been able to build up a worldwide clientele. As of 1995, however, Derwent is engaged in reconsideration of its pricing and marketing practices. It is highly likely that shifts away from prices based on section groupings will take place, particularly when Sections K and L have been priced as if they were the same size.

Relatively brief alerting abstracts are written for new basic references. A feature of the Derwent system is the preparation of expanded titles that aim to capture the heart of the invention. In fact, the titles are sufficiently informative to be used in a series of one chemical and three nonchemical WPI gazettes that appear one week before alerting abstract bulletins. The first appearance of the alerting abstracts is in the on-line WPI database, and can be as early as 4–6 weeks after patent publication for some principal patent offices (United States, EPO, PCT, Germany, United Kingdom); abstracts for Japan take several weeks longer to appear. The alerting abstracts are published in several printed products, including alerting booklets for each CPI section, grouped by country; alerting booklets for each CPI section, classified by technology according to a broad Derwent classification system; and *World Patents Abstracts* (WPA) booklets, covering chemical and, in some instances, nonchemical patents from a number of principal countries and regional offices. The printed abstracts tend to reach subscribers in the United States 3–4 weeks after they appear on-line. Derwent in mid-1995 set a year-end goal of a three-week turnaround time from receipt of patent specifications to completion of document analysis.

In addition to the alerting abstracts, documentation abstracts, formerly called basic abstracts, are produced for all basic chemical patents, except for those from the handful of title-only countries. These documentation abstracts frequently provide a substantial amount of technical details beyond those included in the alerting abstract, although for many Japanese references and others from countries such as Brazil, Hungary, Romania, South Korea, and the CIS, the alerting abstract and documentation abstract may be identical. Documentation abstracts are produced in a format that highlights key features, including claimed matter, uses, and detailed examples; an example is shown in Figure 6, which covers the same invention shown in Figure 2. These abstracts include coding by Derwent's manual code system, which uses a vocabulary of several thousands of keyword-like codes to identify the key aspects of the patent. Documentation abstracts are published in documentation journals covering each CPI section, as well as in profile booklets covering selected segments of polymer and other technology. Microfilm and CD-ROM collections of documentation abstracts are produced for archival purposes.

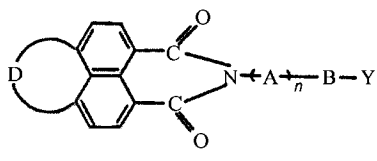
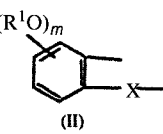
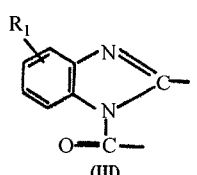
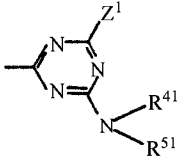
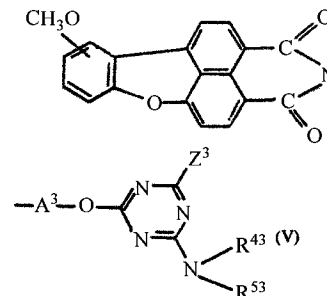
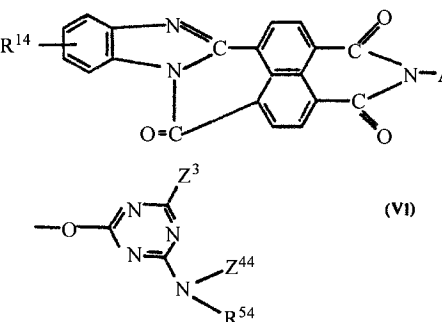
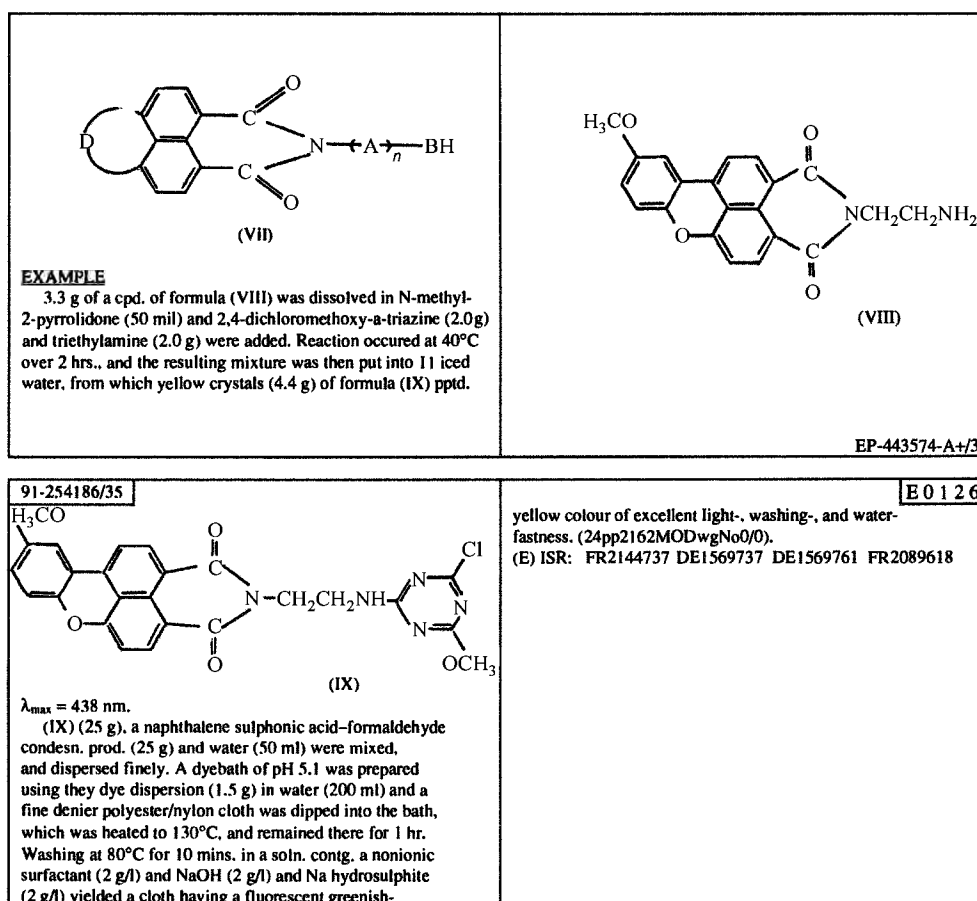
91-254186/35 A60 E23 F06 FARH 21.02.90 HOECHST MITSUBISHI *EP -443-574-A 07.03.90-JP-056212 (+JP-040302) (28.08.91) C09b-57/08 C09b-62/02 New water-insol. naphthalic acid imide dyestuffs - used to dye blends of polyamide or urethane- and polyester or triacetate fibres having good light and washing fastness C91-110342 R(CH DE GB LI)	A(8-E3, 12-S5P) E(25-E1) F(3-F4, 3-F6, 3-F7, 3-F10, 3-F16) E 0 1 2 4
A water-insol. naphthalic acid imide dyestuff of formula (I) is <u>new</u>:  (I) D = a gp. of formula (II) or (III):	 (II)  (III) R¹ = 1-8C (un)subst. alkyl; R² = H, halo, lower alkyl, alkoxy or cyano; X = O or S; m = 0 or 1; A = phenylene, alkylene, aralkylene or C₂H₄OC₂H₄; B = O, S, or N-R³; R³ = H, lower alkyl; Y = 6-membered N-contg. heterocyclic ring having at least one active halo; and n = 0 or 1. EP-443574-A+
USE/ADVANTAGE The dyestuffs are suitable for dyeing fibre mixts. of polyamide or polyurethane fibres, and polyester or triacetate fibres (claimed). The dyestuffs have excellent washing fastness. MORE SPECIFICALLY Y = 6-membered N-contg. heterocyclic ring which may be substd. by 1-4C alkyl, alkoxy, or alkylsulphonyl, cyano, morpholino or amino which may be substd. by alkyl, (alkoxy)alkoxyalkyl or aryl. Pref. Y is a triazine, pyrimidine or pyridazine ring having at least one active halo, esp. a gp. of formula (IV):  (IV)	Z = F or Cl; and R⁴¹ and R⁵¹ = H, 1-8C alkyl which may be substd. by 1-4C alkoxy or 1-4C alkoxy 1-4C alkoxy, aryl or aralkyl, or -NR⁴¹R⁵¹ represents a 6-membered hetero cyclic ring. PREFERRED The dyestuffs are of formulae (V) or (VI):  (V) EP-443574-A+/1
91-254186/35  (VI)	A = 1-3C alkylene or aralkylene; Z³ = F or Cl; R⁴³ and R⁵³ = H, 1-8C alkyl substd. by 1-4C alkoxy or 1-4C alkoxy 1-4C alkoxy; R²⁴ = Cl, methyl or methoxy; R⁴⁴ and R⁵⁴ = as R⁴³ and R⁵³; and A⁴ = as A³. PREPARATION (NOT CLAIMED) (I) can be prepd. by heating a water-insoluble cpd. of formula (VII) with a cpd. of formula Hal-Y (Hal = active halogen atom), which is present at a level of 1-1.2 molar times that of (VII). Reaction occurs in an organic solvent e.g. acetone, toluene or DMSO, and in the presence of an acidic binder (1-2 molar times the level of the water insoluble cpd.) such as a tert. amine or an inorganic base, at a temp. of 0-90°C over 1/2-5 hrs. EP-443574-A+/2

Fig. 6. Representative Derwent documentation abstract.

Courtesy of Derwent Information Ltd.



The manual code system was originally created to enable the searching of classified sets of documentation abstracts grouped by manual code. The system was analogous to the traditional method of searching classified sets of full patent specifications, and has proved to be an effective search method over the years; however, some organizations that once relied on manual code searches have stopped maintaining the card sets because of the cost and space involved in acquiring and filing the cards. Manual codes can also be used for searching in the printed documentation abstract journals. Assigned to all chemical and electrical patents except some title-only references, manual codes are included with the alerting abstracts and bibliographic data when references are added to the WPI database, and remain a valuable search parameter in on-line searching.

An important advance in the on-line WPI database is the inclusion of representative drawings from patents. These drawings are particularly important in conveying the meaning of mechanical, engineering, and electrical inventions, but are also important in elucidating the chemical structures involved in chemical inventions. The drawings, which can be read on-line and printed on- or off-line, cover chemical patents from 1992 onward and nonchemical patents since 1988.

The manual code system covers the entire CPI. In addition, specialized deep coding and indexing systems exist for Sections A, B, C, and E. Section A features the Plasdoc code, introduced at the start of polymer coverage in 1966, and enhanced on a number of occasions since then (8). During 1993, Derwent introduced an ambitious new polymer indexing system, that has a greatly expanded indexing vocabulary and the capability to pinpoint contextual relationships among substances and concepts contained in a given patent via a three-tiered linking system (9). In Sections B, C, and E, a chemical fragmentation code has existed since the start of pharmaceutical coverage in 1963, and it too has been enhanced frequently over the years (10). In 1987, Derwent introduced a topological structure coding and retrieval system aimed at coping more accurately with Markush or generic structures in patents. To supplement their long-standing series of abstract bulletins, Derwent has produced a number of repackaged sets of abstracts. Some of these are targeted at specific industries, such as the automobile industry. Others are created to match the interests of individual companies or groups of companies. Formats of these specialized bulletins can vary considerably. They may include alerting abstracts, documentation abstracts, or even abstracts from other sources; they may be produced as paper copy or in electronic form. This particular Derwent activity is likely to continue in a state of flux as new opportunities are identified and pursued. *Patents Preview* and *World Drug Alerts*, the latter encompasses journal literature and conference proceedings as well as patent information, represent additional Derwent approaches to satisfying the needs for rapid information dissemination in the pharmaceutical industry.

Besides the worldwide WPI database, Derwent provides on the ORBIT system the USPatents database, a bibliographic file of patent front page and claim information for U.S. patents since 1971. Derwent also produces a biotechnology database, GENESEQ, that indexes sequence structures of proteins or nucleic acids disclosed specifically or generically in patents. This database is searchable with special sequence software on the IntelliGenetics system, and is a new addition to STN's database catalog.

Abstracts can serve as a pointer to patents of potential interest, and in some cases may provide sufficient information to judge the relevance of a patent, but there is no true substitute for the examination of complete patent specifications, particularly when legal decisions must be made. An aspect of the Derwent service since its early days has been the provision of microfilm and, in selected instances, paper copies of patents covered by the system. In most instances, the basic patent and key English-language equivalents have been included. English abridgments of Japanese patents were included until the early 1970s, when the volume of Japanese documents produced under the then-new Japanese patent law made this impractical. Complete patent specifications have been provided in groupings by Derwent section. Derwent in the past offered country groupings as well, but transferred that operation to an associated company, Research Publications Inc., which has subsequently been integrated into the Derwent operation as Derwent Direct. Derwent has also offered a document delivery service for worldwide patent copies as well as a translation service for non-English patents. Document delivery services have also been offered by Rapid Patent International, which also became a part of the Derwent organization. The responsibilities of these two Derwent affiliates are evolving and expanding.

Derwent has developed, either on its own or through contractors, a number of computer aids to information processing. The TOPFRAG series of programs, available as MARKUSH TOPFRAG, aids users in searching chemical structure information including the generation of search strategies for the WPI MARKUSH database and for fragmentation code searching of the CPI. MARKUSH TOPFRAG has also been embedded in STN Express software as an aid in searching structures on Derwent files in the STN system. PILOT is a program for assisting searchers with strategies under the new polymer indexing system, and in its next upgrade is expected to provide searching strategies for the back-file Plasdoc code. Derwent InfoView is a program for the manipulation and analysis of patent statistics, replacing the earlier PATSTAT PLUS.

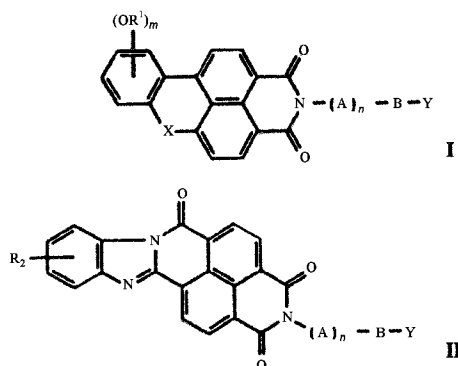
A new initiative introduced by Derwent during 1995 is the Patents Citation Index, an on-line database of patent citations that includes both examiners' citations and patentees' citations to prior art from patent specifications. When given a known invention of interest, as represented by a patent family, the database can identify any patent against which it has been cited, as well as earlier patents cited by any member of that family. Limited citation

searching capability has long been available in a number of databases, including WPI and USPatents, but Derwent's new file greatly increases the ability to carry out citation searches. Finally, Derwent provides a paid search service that carries out searches of Derwent and other databases (qv).

Chemical Abstracts Service. The Chemical Abstracts Service (CAS), a division of the American Chemical Society, has produced *Chemical Abstracts* (CA) since 1907. Since the demise of *Chemisches Zentralblatt* and *British Chemical Abstracts*, CA has been the preeminent medium for documenting new publications in the field of chemistry and chemical engineering. CA documents chemical publications of all types. It is not a patent database per se, but its patent component is larger than most databases devoted entirely to patents. Thus, for example, the number of patent references in CA for the years 1991–1993 ranged from 95,500–99,400 per year.

Derwent products have always been targeted at those in and near the legal profession, and the company itself has always attempted to emphasize in its treatment the fact that patents are legal documents. CA, on the other hand, has the mission of documenting chemistry for chemists and chemical engineers. Therefore, CA abstracts of patents have emphasized what was actually done in examples that provide hard data, and have avoided discussing the purpose or scope of a patent or prophetic paper examples. The CA abstract for the same patent whose Derwent abstract appeared as Figure 6 is shown in Figure 7. Until about 1980, CA's abstracts and indexes concentrated only on examples, not claims; however, from about 1980 onward, CA has been indexing substances covered explicitly in claims even if these are not described in examples with hard data. Another significant change since 1988 has been the structural indexing of Markush structures from patents in the MARPAT database.

116: 131162r Fluorescent reactive disperse naphthalimide dyes. Himeno, Kiyoshi; Hihara, Toshio (Hoechst Mitsubishi Kasei Co., Ltd.) **Eur. Pat. Appl. EP 443,574** (Cl. C09B62/022), 28 Aug 1991, JP Appl. 90/40,302, 21 Feb 1990; 24 pp. The dyes, esp. useful for blends of polyamide or polyurethane fibers with polyester or triacetate fibers, have the general structure I or II [A = C₆H₄, alkylene, aralkylene, CH₂CH₂OCH₂CH₂; B = O, S, NR³; R¹ =



(un)substituted C₁₋₈-alkyl; R² = H, halogen, alkyl, alkoxy, CN; R³ = H, alkyl; X = O, S; Y = active halogen-bearing N-heterocyclyl; m, n = 0, 1]. Thus, the appropriate N-(2-aminoethyl)naphthalimide was condensed with 2,4-dichloro-6-methoxy-s-triazine to give I [A = CH₂CH₂, B = NH, R¹ = Me (OR¹ para to X), X = O, Y = 4-chloro-6-methoxy-s-triazin-2-yl, m = n = 1], λ_{max} 438 nm in DMF, giving light- and washfast fluorescent greenish yellow shades on an 8:2 polyester-nylon blend.

Fig. 7. Representative abstract from *Chemical Abstracts*.

Courtesy of the American Chemical Society.

Just as Derwent has broadened its country coverage, so has CAS. During the 1960s, CAS policy was to cover all chemical patents for only a handful of countries; for other countries, only patents to nationals were covered. The group of countries changed from time to time, and at times only the United States, the United Kingdom, and Germany were fully covered. French coverage was theoretically also complete, but in practice less than that. That was the era before the widespread introduction of the practice of publishing all patent applications. The United States, the United Kingdom, and Germany published only those applications determined to be patentable, and one consequence of the CAS coverage policy was that many inventions that appeared only in countries where coverage was limited to nationals did not get abstracted or indexed. Beginning around 1968, the number of countries afforded full coverage for chemical patents has increased, so that CA's coverage of chemical patents is far more complete in the 1990s. Coverage now extends to 26 countries as well as the EPO and PCT. Only eight of these countries are limited to nationals-only coverage.

There is an important difference, however, between the CAS and the Derwent treatment of members of a patent family. Derwent abstracts the first member of a family that it sees, then adds bibliographic data for all equivalent patents to the record, so that the record for a given invention in the WPI database can be accessed by bibliographic information on any member of the family. Although CAS also abstracts the first member it sees of a family, it enters subsequent equivalents into only the printed CA patent index, not the on-line CA or CASearch databases. Thus a searcher looking for a given patent number in CA will find it only if it was the member of a family that happened to be covered first. Priority application details common to a basic patent and its equivalents are thus more reliable than patent numbers as an access point to the CA database, although indirect access based on patent number can also be achieved via the cross-file techniques of various on-line systems.

EPIDOS (Formerly INPADOC). The International Patent Documentation Center (INPADOC) was created as a result of agreements reached in 1972 between WIPO and the government of Austria. It reflected the desire of many in the intellectual property community to have an authoritative and noncommercial repository and dissemination center for bibliographic information on patents. The INPADOC operation has now become part of the division of the European Patent Office known as the European Patent Information and Documentation Systems (EPIDOS). Information is obtained directly from national and international patent offices, which in the 1990s number around 60 and include more countries than any other patent information service. Title, inventor, patentee, classification, as well as priority, application, and publication details are all included for different stages of publication, including in some cases unexamined, examined, and granted patents, and even unpublished applications for some countries. EPIDOS ties together members of extended convention patent families, and in some instances identifies intellectual families as well. Legal status subsequent to publication is also collected for 16 authorities.

EPIDOS issues printed and microfiche compilations of its data; in addition, its database can be searched on its own computer or on several on-line host systems. In general, EPIDOS provides the most complete patent family information of any service, although Derwent tends to include more information on intellectual (nonconvention) families, whereas the French Patent Office's EDOC file on the Questel system includes information on

granted Japanese patents that is not readily available elsewhere. Also available from the EPO is the EPIDOS Register file, which is unique in providing information on the course of prosecution of European patent applications. An important EPIDOS component is the CAPRI system, under which patent documents going back to 1920 and beyond have been reclassified according to the IPC system. EPIDOS also carries out individual searches and provides patent copies as well as information on Japanese patents, including English-language abstracts and searches of the Japanese-language PATOLIS database.

An important new service from EPIDOS is the series of ESPACE CD-ROM products, providing document delivery of full patent specifications from the EPO, PCT, and a lengthening list of individual countries. Approximately 1000 full specifications can be contained on an individual CD-ROM for printed documents such as European or U.S. patents. PCT applications, typed with wide spacing between lines, require more storage space. Other ESPACE CD-ROMs provide in searchable form the front pages of European and PCT patents, the *EPO Bulletin*, legal decisions, and other information of interest with regard to intellectual property. New CD-ROM products continue to appear regularly.

IFI/Plenum Data Corp. IFI Plenum's predecessor company, Information for Industry (IFI), began indexing U.S. patents by its Uniterm system in 1955. Coverage was eventually extended back to 1950. In 1972, the Uniterm system was complemented by a more powerful retrieval system based on a merger with an indexing system developed by Du Pont and acquired by IFI (11,12). The latter system, called the Comprehensive Data Base, is available only to subscriber organizations. With the advent of on-line databases, these chemical indexing systems were augmented by bibliographic information, including bibliographic data for nonchemical patents going back to 1963.

IFI has never been a patent abstracting organization, although it does provide subscribers with sets of bibliographic, abstract, and claim information selected from U.S. PTO tapes to match the subscribers' interest. Besides its on-line databases, IFI produces magnetic tape versions of the databases which some users choose to run in-house. Other IFI patent products include the *IFI Assignee Index* to U.S. patents, produced on a quarterly basis and thus available well in advance of the U.S. PTO's published index, and the *Patent Intelligence and Technology Report*, an annual listing of patents to all organizations receiving at least 10 U.S. patents during the preceding year. These reports show the total number of patents, and a breakdown by U.S. patent class. One shortcoming of these reports, as well as of patent count lists issued by the U.S. PTO itself, is the fact that they do not aggregate the patents for those industrial organizations that choose to have their patents assigned to multiple subsidiaries or divisions. Some corporations receive patents through 20 or more subdivisions, whereas others use a single patenting entity. Comparisons made on this basis can thus be quite distorted.

L'Institut National de la Propri  t   Industrielle (INPI). The French Patent Office (INPI), is a principal provider of patent and trademark databases, all of them accessible on Questel. FPAT, EPAT, and PCTPAT have full bibliographic data, abstracts, and claim text for French, European, and PCT patent documents, including information about changes in the status of the applications after their original publication. The text of the European Patent Classification system (ECLA) is searchable in the ECLATX file, which is updated monthly with changes made to accommodate changes in technology reflected in patent applications. EDOC contains the ECLA codes assigned to the documents in the EPO search files. The patents in EDOC are organized into patent families with a single set of ECLA codes, thus serving as a source of patent family information as well as a means for searching international patents with a uniformly assigned classification system. PHARMSEARCH is a structure-searchable file of pharmaceutical patents from France, the EPO, the PCT, the United States, the United Kingdom, and Germany (13). The database is searchable through the Markush DARC system and has extensive indexing of pharmaceutical concepts, abstracts, and displayable images. The original database was INPI's pharmaceutical structure search file, and patents in the file are indexed retrospectively; French special medical patents published between 1961 and 1973 have been indexed with Markush DARC, thus adding structure-search capability to a small segment of the early patent literature.

American Petroleum Institute. The American Petroleum Institute's Central Abstracting and Information Services are a unique example of the creation and molding of a series of information resources, including a patent database, by a group of companies with similar interests (14). During the 1950s, it was common in the petroleum and petrochemical industry for individual companies to have in-house abstracting and indexing groups. As costs of such operations increased, these companies turned to the American Petroleum Institute (API) to provide a vehicle for centralized production of information services. Literature coverage and a specialized bulletin on Soviet literature began in the 1950s, and a printed patent abstract bulletin began in 1961.

During the early 1960s, cooperative efforts under the API resulted in the development of a system for indexing these bulletins for searching by computer. The heart of the indexing system was a thesaurus developed by a study of a year's worth of published literature in the fields of interest. The resulting APIPAT and APILIT databases were launched in 1964, and were among the first databases to go on-line in the mid-1970s (15).

In 1972, the API reached an agreement with Derwent to use repackaged Derwent alerting abstracts for its printed patent bulletins and to do its patent indexing from Derwent documentation abstracts. This enabled API to discontinue patent abstracting for the most part, and the documentation abstracts provided richer material for indexing than did the relatively brief API abstracts. Cross-referencing from the APIPAT database to WPI enabled on-line searchers to move from hits in the APIPAT database to the corresponding WPI references, including their complete patent families. Ultimately, the APIPAT and WPI databases were merged on the ORBIT system, which enables a searcher to combine API and Derwent retrieval parameters in a search (16). APIPAT and WPI remain separate databases on the DIALOG and STN systems.

Perhaps the most notable aspect of the history of the API operation is that it has been shaped at every step by those who use the system. Created by information specialists within the petroleum industry, it has been governed by a technical information committee made up of company representatives, and guided by a series of industry task forces, which have modified as needed the indexing thesaurus, subject selection guidelines, and selection rules for countries in patent coverage, journals, and other sources in the nonpatent literature.

Types of Patent Information Searches

There are many different reasons to search for information about or related to patents. The methods, sources, and techniques vary widely, depending on the purpose and the complexity of the individual situation. Searches can cost from \$5 to well over \$5000. It is essential when searching for patent information to gear the strategy to the task at hand. An ill-conceived computer search strategy can produce mountains of expensive output that can require a huge outlay to analyze. An inadequate search strategy can lead to even greater costs if the result is patent infringement. Anyone performing patent searches must have a sound understanding of the costs and benefits that may be involved.

Novelty Searching. At the heart of the patenting process is the novelty or patentability search (17). A novelty search should be carried out by an inventor or an inventor's representative before a patent application is drafted in order to help ascertain whether the invention is indeed patentable and, if so, what its limits might be. A novelty search is carried out by a patent office examiner to make a decision on the patentability of an invention. Although there may be in some instances a temptation for an inventor to omit or ignore the novelty search and rely on the examiner's work, the cost of making a patent application is considerable. Further, the failure to be aware of relevant prior art when filing an original application can place limitations on the applicant's ability to reshape the application by amendments. A novelty search is normally focused sharply on the specific details of the invention, but broader searches that provide a context for the invention in relation to the state of the art can be justified when an invention gives promise of having wide application and high value.

A valid patent covering any claimed invention can be obtained only if the wording of the claims defines an invention that has never been used or described before the filing of the patent application, and that is not an obvious variation of something that has been described in the prior art. However, the standard for prior art references that may be brought to bear against a patent differs from country to country. Thus, for most countries, a standard of absolute novelty applies, ie, references can come from anywhere in the world and can have been published at any time prior to the priority filing date of the patent application. For other countries, which are far fewer than in the past, only references that reside physically within the country are considered. Exceptions are made in some countries for public disclosures by the inventors during a short grace period prior to the filing of the application claiming the invention, or for public use of the invention for experimental purposes. In the United States, which grants patents to the first inventor rather than the first to file a patent application, a description of the invention published less than a year before the application was filed may be discounted as prior art if the applicant can prove that the invention was made prior to the publication date. Standards for judging whether an invention is unpatentably obvious also differ from country to country. Most countries will not grant a patent covering a claimed invention that is similar enough to one described in a previous publication so that a person working in the field would consider the differences trivial. Thus, a publication showing a process using ethanol as a solvent would prevent in most countries the grant of a patent on a similar process that uses methanol. Most countries consider an invention to be unpatentable when information from two separate publications can be combined to match the claimed invention. Standards for judging whether an invention is

unobvious or has an inventive step over the prior art are based on precedents set in published decisions of the patent office or national courts.

Traditionally, novelty searches have been performed by leafing through stacks of patents in those divisions of a classification that seem best to categorize the invention. This sort of searching is facilitated greatly by the front page of the modern patent specification, which can often show the searcher at a glance whether or not the patent is likely to be relevant. It requires, however, that the searcher have access to appropriate sets of classified patents and that the integrity of the collection of patents is maintained. Further, it excludes from consideration those patents that might be relevant to the search but which, for hierarchical or other reasons, are classified elsewhere. For example, two patents may be closely related. One may be classified under the product and not cross-referenced to the process, the other classified under the process but not cross-referenced under the product. As a result, the earlier patent would not have been considered when the later one was granted. This is because an examiner may cross-classify a patent, which has become more and more the common practice, but legally the examiner is required only to classify a patent in the most relevant class. Thus, searching based solely on classification risks the omission of useful references.

Searching of one or more on-line databases is a technique increasingly used in novelty studies. The use of such databases enables the searcher to combine indexing parameters, including national and international classifications; natural language words in the full text of patents, in their claims, or in abstracts supplied by inventor and by professional documentation services; and indexing systems of various sorts. Because the various patent databases have strengths and weaknesses that complement each other, the use of multiple databases is thus prudent, and is facilitated by multiframe and cross-frame techniques provided by the various on-line hosts.

On-line searches carried out in this way can provide impressive recall of potentially relevant documents, but depending on the search strategy used the results can be quite dilute; they can include many references that are not relevant. The evaluation of the results of on-line searches can be more difficult than the evaluation of a hand search through classified patent specifications. The searcher in general obtains a computer printout, which may be quite lengthy and many of whose listings may provide insufficient information to determine relevance. Output from a simple bibliographic database such as INPADOC provides no subject information beyond a title, often uselessly brief, and patent classes. Abstracts from CA or WPI can be more helpful in determining relevance, but can still leave much open to question. The recent addition of drawings and some structural information to on-line WPI records is certainly a step forward, but falls short of the full structural information needed for many decisions. Derwent documentation abstracts and their rich information content would be useful as a primary search output, but as of 1995 are not directly available from searches of on-line databases. Potential routes to such availability are under consideration or development. Derwent is producing collections of documentation abstracts on CD-ROM, and it is possible to download the results of an on-line search and use those results directly to print copies of documentation abstracts from CD-ROM; however, systems for doing so are in need of further development. Derwent is also exploring mechanisms for supplying users with sets of documentation abstracts retrieved in on-line searches. The direct availability of documentation abstracts on-line is yet another possibility.

Documentation abstracts, useful as they are, may also prove inadequate for final decisions, which would make it necessary for the searcher to obtain and examine copies of full patent specifications. Although the increasing availability of sets of full specifications on CD-ROM at modest cost is making it easier for searchers to have in their collections copies of those specifications they might need to consult, the necessity to go beyond computer output, whether to abstracts or full specifications, is still one of the bottlenecks of computer searching, and therefore an area in which significant new developments are hoped for.

Novelty searches are not necessarily limited to patent information. The anticipation of a purportedly novel idea can occur in journals, books, magazines, etc. Thus, the potential scope for a novelty search is essentially infinite, and one of the challenges to the searcher is to devise an effective strategy whose cost is commensurate with the potential value of the invention.

Infringement Searching. An individual or organization found to be infringing the patent rights of others is subject to penalties that can be extremely costly. It is essential for anyone contemplating a commercial venture that is technology-dependent to find out first whether or not the proposed venture falls within the area covered by adversely held patents.

Whereas the potential field of search for the novelty search is essentially limitless, there are certain limits that can be placed on an infringement search. References, either new or old, relevant to a novelty search can appear in any medium, patent or nonpatent, anywhere in the world; they may be found in claims or in disclosures anywhere in a patent specification. By contrast, an infringement search can be limited to the content of the claims of patents, and only to the country or countries in which manufacturing, sale, or use of the invention is contemplated. Only patents that are in force or that are potentially in force need to be considered. Patents that have expired, that have been invalidated, or that have lapsed because of failure to pay maintenance fees can be excluded from consideration. A searcher must be alert, however, to patents that are potentially in force. Thus, if the United States is the country of interest, attention should be given to patent cases that have been published in other countries and are likely to be pending in the United States. If the intention is to operate in one or more countries outside of the United States, consideration must be given to published applications that have not passed examination but that might be granted in the future. Also, a patent on a given substance might exist without composition of matter claims but include process claims that in themselves are not of concern. The searcher must also consider the possibility that a divisional patent with composition claims might yet be issued.

Since the exact language of claims is vital to matters of infringement, the search of full patent specifications remains the most reliable method of infringement searching. However, this is no longer restricted to hand searching of hard copy documents. The full claims text of all U.S. patents is available in several on-line databases covering a time span longer than the life of a U.S. patent. A full-text file of European patents has been announced by DIALOG and by STN, but is not yet available, and its future seems uncertain. In the meantime, there are exemplary European, German, French, and PCT claims on-line, but infringement takes place when any claim, not just the representative claim, is infringed; there are also many countries whose patent claims are not yet searchable on-line. On-line databases are sometimes used for infringement searches by carrying out careful searches of parameters other than the claim language. However, reliance on computerized databases lacking full claims text for infringement searches involves compromises; at the very least the searcher must obtain the full claims text, including any associated drawings or chemical structure diagrams, of all patents of potential interest that are disclosed by the computer search.

Validity and Opposition Searches. Given the identification of a patent that presents a potential infringement risk, an individual or organization may choose to obtain a validity study in the hope that references can be located which show that an invention was either anticipated or obvious, and that the patent should not have been granted. As was the case with novelty searches, the potential scope for a validity search is broad. It can include both patent and nonpatent literature. In particular, a disclosure in a patent specification not closely tied with that patent's claims can often be useful in invalidating a patent. Since such disclosures are typically not reflected in the classification of the patent, which is tied to the claims, classified sets of patents are not necessarily effective for validity searching. Deep-indexed databases on the other hand can be useful, and full-text patent databases also promise great utility in validity searching. The stakes involved in gaining freedom from blocking patents can be substantial; the cost and effort expended in a validity search can be correspondingly large.

Closely akin to validity searching is searching for the purpose of opposition. Long a factor outside the United States, this technique is becoming increasingly important in the United States as companies engage more and more in worldwide operations. In most countries outside the United States, when the examiner has been satisfied that an invention is patentable, the patent specification is published either as an examined application or as a granted patent, and third parties are given a limited period during which to oppose or challenge the patent. Validity and opposition searches have the same requirements as novelty searches, ie, any reference that would render an invention unpatentable under national laws is relevant in an opposition search. Unlike novelty searches, however, opposition searches are performed long enough after the filing of the patent application so that all of the prior art published before the date of filing is made available for searching.

State-of-the-Art Searches. State-of-the-art searches are typically carried out when research in a newer area is to begin in order to identify what has previously been done, what is known, and where fruitful opportunities might be found. Typically, a state-of-the-art search is broad and general, although tighter and more focused follow-up searches are often carried out once the areas of potential interest are identified. Detailed state-of-the-art searches are not the norm, but it is quite common for organizations to prepare and maintain such detailed studies in subject areas of great importance and significant commercial or research interests.

Alerting Searches. Various means are available for keeping up with the latest in patents, and it can be effective to use a combination of these

methods. Thus, computer profiles created to represent individual, group, or organization interests can be run against databases as they are updated, and specific searches can also be run against these databases. There are databases that are updated promptly. INPI's FPAT file of French patents is updated on the date the patents and applications are published. EPAT, which was formerly updated on the publication date, is currently subject to a two-day delay. The U.S. patent files are typically updated in less than a week. Limited U.S. patent data are available from a United States PTO electronic bulletin board on the day of issue. Derwent WPI data, using value-added access points, is on-line roughly 5–6 weeks after patent issuance from principal patent offices; this time is expected to be cut sharply. Other value-added databases tend to be slower. CA is quite variable in its timing; some patents can be covered rapidly, others lag substantially. The CAPlus database on STN helps in providing an advance look at unpolished abstracts and indexing slated for the CA file. The MARPAT Previews file gives early access to Markush structure information from patents.

Other printed and on-line products are aimed at alerting in targeted areas, especially in pharmaceuticals. Current Drugs Ltd. produces short abstracts of U.S. patents and British, PCT, and European patent applications within two weeks of issue for the printed *Current Patents Fast Alert* service and the corresponding on-line file. These abstracts emphasize the pharmacological and pharmaceutical aspects of the patent disclosures rather than the legal content of the patent. Derwent publishes patent abstracts from a larger group of countries in a similar format in *Patents Preview* and *World Drug Alerts*, which feature the novel aspects of the inventions as well as their pharmaceutical utility. These abstracts are different in format from CPI abstracts and are distributed before the corresponding CPI record is created for printed or on-line access. When abstracts are prepared within days or weeks of patent issuance, the speed of alerting services hinges more on delivery routes than on production rates. These services are also available in electronic form for access within subscriber organizations, and updates can now be transmitted electronically, thus avoiding mail delays.

Traditional paging through patent office gazettes and printed abstract bulletins still serves a useful purpose in patent alerting. It can be difficult to frame a query for a computer search on all the subject matter that might be of interest to an organization. The human mind can spot unanticipated material and relate it to interests, something presently beyond the power of the computer.

Most new patent cases of interest are published by at least one of the U.S., European, or Japanese patent offices, and WIPO (PCT). Japan presents problems for those not able to read Japanese, but the U.S. *Official Gazette* (with representative claims) and *PCT Gazette* (with English-language abstracts) can be in one's hands within a week of patent publication. Similar timing is available for the *European Patent Office Bulletin* which contains trilingual titles and the on-line EPAT file and various CD-ROM products. A highly effective alerting program can be developed from a combination of these methods.

Family and Equivalent Searches. A wide range of inquiries fall under the category of patent family searches. It may be desirable to find an equivalent to a known patent in a given language, typically but not necessarily English. It may be necessary to find whether an invention is protected in a given country. It may be desirable to estimate a patentee's interest in an invention on the basis of how broadly it has been filed, or to know in detail all the countries in which an invention has been patented, including the legal status of each. Or it may be necessary to trace the entirety of a complex extended family, replete with divisionals, continuations, continuations-in-part, and multicountry equivalents at one or more stages. All of these tasks have become relatively simple because of the efforts of Derwent, EPIDOS, and INPI. Derwent's WPI database covers 40 patenting authorities in 1995, and provides information on multiple stages of publication in many of them. It identifies many intellectual patent families, and provides data links that make tracing the web of extended families possible. Family information goes back to 1970 for chemistry and several years earlier than that for pharmaceuticals, agriculturals, and polymers. The ORBIT version merged with APIPAT also includes family information from the 1960s relating to petroleum and petrochemicals.

EPIDOS in the INPADOC database covers even more countries and stages of patent publication, and includes patent status information for 16 authorities. Coverage for most major countries goes back to 1968 or 1973; starting times for other countries vary. Family searches in the INPADOC database identify extended families. EPIDOS identifies some intellectual families, though fewer than Derwent; however, it is often possible in searching INPADOC to identify nonconvention family relationships. Family searches in INPADOC are more expensive than family searches in WPI, so that those looking for simple family information, such as an English equivalent or an indication of the breadth of filing, normally use WPI for this type of information.

EDOC, available on the Questel host from INPI, is unique among non-Japanese language databases in including information on C-stage Japanese patents, ie, those that have successfully weathered the pregrant opposition period and been sealed as patents under pre-1966 patent law. It also contains some information on patent family relationships from the period long before the advent of patent family databases.

An often forgotten source for some patent family information is the CLAIMS database. Although its direct coverage is limited to U.S. patents, it includes limited patent family information up to 1979, for Belgium, France, Germany, the Netherlands, and the United Kingdom. Some of this information from the pre-1970 era is found in no other on-line database. An earlier discussion of patent family databases has recently been updated (18).

The *Chemical Abstracts* database does not have an on-line family capability, but does publish a patent index, with family data obtained from INPADOC. Printed patent indexes have been including family information since the 1960s, but the number of countries covered before 1970 was limited.

Citation Searching. In the scholarly literature, authors cite earlier publications that relate to the work being reported, thus a subject relationship exists between the citing and cited literature. This relationship has formed the basis for the *Science Citation Index* and related products, developed by the Institute for Scientific Information. Known as Scisearch in its on-line version, the *Science Citation Index* has become an important information retrieval tool in the second half of the twentieth century. It has been used for straightforward subject searching, in which mode it complements traditional indexed databases and indexes. It has also become a popular tool for bibliometric studies of various sorts, such as attempts to measure the relative impact of research carried out by different individuals or organizations, or the relative impact of publications in different journals.

Citations appear in patents as well as in the journal literature, and it has been proposed that they, too, should become searchable to provide different types of useful information, such as research trends and estimates of the effectiveness of research organizations. Elaborate techniques for using citation data for such analyses have been developed (19,20). However, citations found in patents can differ in a number of ways from literature citations and these differences can strongly color search results (21,22).

At least three types of citations in patents can be identified: inventors' citations found in the patent specification, examiners' citations found on issued U.S. patents, and examiners' citations found on published applications and granted patents from other countries. A patent inventor cites prior art in order to distance the invention from that art, rather than to show a close relationship. Whereas scientific researchers may want to show how closely they have built on what went before, for an inventor that can suggest anticipation or at least obviousness. Thus, citations within a patent typically try to demonstrate the inadequacies of prior inventions and the uniqueness of the patentee's own work. References tied by this type of citation can be useful in developing a picture of the state of the art, but often show sharply differing technologies.

An examiner's citation found on a U.S. patent should show art that is related to the invention at hand, but which did not anticipate that invention. If the invention had been anticipated, the patent would not have been issued. When a patent is an improvement patent on an earlier invention, examiner's citations typically show the fundamental invention and represent an analogue to the traditional literature citation. In addition, examiners' citations often are used to show general background and the state of the art. However, there seems to be an increasing tendency among U.S. examiners to cite dozens, even more than 100 earlier references, even though long citation lists dilute the meaning of citations. Furthermore, some examiners appear to have personal favorites that they cite whenever a given subject area comes up, regardless of how closely the technologies match. Both of these factors diminish the significance of some citation searches of U.S. patents.

A distinct difference between examiners' citations on granted U.S. patents and those on published patent applications is that the latter can indeed represent direct anticipation. Thus they represent a close subject relationship to the document in question. An important factor in the citations on EPO and PCT applications is that they are categorized by the examiner with regard to their relevance: documents of particular relevance in themselves, documents of particular relevance in combination with some other document(s), and documents defining the general state of the art but of no particular relevance in themselves. Clearly not all citations have the same value.

Citation searching of patents offers a perspective different from either traditional class searching or traditional subject index searching. A citation search on known fundamental patents can lead directly to improvement patents, even when those patents are so new that they have not yet been indexed. This technique can be especially effective when working in an unfamiliar area, or one which is difficult to index.

The availability of citation searching tools is on the increase. A citation database for U.S. patents was first built by Search Check, Inc., which collected all U.S. patent citations back to their first appearance on patent copies in 1947. This database was subsequently made available for on-line

searching as the CLAIMS-Citation database. Lower priced citation searching of U.S. patents from 1971 onward has been available since the advent of the USPatents files on ORBIT, and the examiners' citations in EP and WO patent documents have been available in the WPI database. Citations can also be searched in some of the individual country patent databases. In 1995, Derwent introduced the Patents Citation Index covering inventors' and examiners' citations from 16 countries, and enabling searches to be carried out for citing patents as well as for cited patent and nonpatent references. The availability of this tool sharply increases capabilities for citation searching, and promises better exploitation of an intriguing type of patent information.

Business-Related Searches. Many searches of business-related questions can be answered by searches of patent information. For example, an organization may wish to study the patent assets of competitors in a technical area or to evaluate similarities and differences in approach and strategy between its own and other organizations. Statistical analyses based on citations and other data may be desired. Searches may also be desired to identify candidates for joint ventures or for acquisitions or divestitures, or to clarify the relationships of corporate segments. Knowledge of the technology behind new product or process announcements by competitors, or the technology being offered for license or purchase by an individual or small organization, may likewise be needed. Searches of patent databases are invaluable in answering all of these types of questions.

On-Line Database Searching Methods

Coordinate Indexing and Boolean Logic. Three methods of indexing have been prominent in the chemical literature in recent times. The first, articulated indexing, has been used in printed *Chemical Abstracts* subject indexes from their earliest days until well into the 1990s. A number of important concepts are identified as permissible index entries, including specific compounds, material types, reactions, and processes. One or more modifying statements follow each basic index entry. Thus, eg,

Hydrocarbons, pyrolysis of, in plasmas

A second type of index, the keyword-in-context (KWIC) index, arose during the early days of computer processing. The same entry would appear in a KWIC index as follows:

PYROLYSIS OF HYDROCARBONS IN PLASMAS

The third type of index is the coordinate index, in which all of the individually indexable concepts of a document are posted to the record for that document. Entries in coordinate indexes may be based on groups of words, single words, or codes. Coordinate indexes became significant during the 1950s and 1960s in such tools as the Uniterm Index to U.S. Chemical Patents, as well as in personal information tools such as optical coincidence cards and edge-notched cards. The aforementioned index entry would produce three indexing terms:

HYDROCARBONS, PYROLYSIS, and PLASMAS

In the earliest days of on-line databases, all three indexing types collapsed into the third. Using older manual tools, it was difficult to coordinate more than two or three concepts, but the computer made that easier. Each concept in a search can be represented by a string of synonyms or alternatives, and searching can be done for two such parameters or more. Thus, Boolean logic expressions can easily be constructed as follows:

(A1 OR A2 OR A3) AND (B1 OR B2 OR B3) AND
(C1 OR C2 OR C3 OR C4) AND ...

However, this advance has an important shortcoming: the lack of context. More than one idea is expressed in a document; a patent on oxidation catalysts, for example, could include examples of the oxidation of methanol to formaldehyde and of 2-propanol to acetone. A simple coordinate search for conversion of methanol to acetone would retrieve such a document from a file that provides no context.

A number of methods have been developed to introduce context to on-line databases, enabling searches to be refined to minimized false retrieval. One of the earliest techniques is proximity searching, in which two words are required to be adjacent, or within a limited distance from each other in text. The assignment of roles to chemical substances is a method of precoordinating concepts. A substance can be identified as a reactant, as a product, and in some systems in a number of additional roles. For example, by searching for documents in which formaldehyde is a product, documents in which it is a reactant, or in which it undergoes no reaction, are thus eliminated.

Another source of context comes from links between index concepts. In a database that describes chemical compounds in terms of their fragments, it is important that those fragments are tied together, and that the fragments of compound A are tied separately from the fragments of compound B. IFI Plenum's Uniterm Index has a simple method of fragmenting chemical structures, but one of its shortcomings is the fact that there is no linking of the fragments. Chemical fragments are linked in IFI's Comprehensive Data Base, and this is one of the reasons for CDB's ability to outperform Uniterm. Linking logic has enabled the articulated index terms of CA to come back into their own and restore their original context. One of the most interesting developments in on-line database indexing has been the introduction of three levels of linking in the revised polymer indexing system introduced by Derwent late in 1993. This method improves the ability of the searcher to look in an overall system for subsystems that bear a given relationship to each other.

Subject-Based Retrieval Parameters. There are numerous means by which the subject content of a patent can be expressed, and which a searcher can use in developing a search strategy. Different databases offer differing subsets of these means. Effective strategies should in general not be limited to a single type of retrieval parameter; rather, they should be built from different parameters and modified as needed to provide the strategy best fitted to the subject at hand.

Patent titles are usually short, and sometimes extremely uninformative. Patents in the latter category include titles such as "Chemical Product" and "Process." Derwent rewrites titles to make them more informative, and IFI Plenum augments many titles. Words included in patent titles normally are highly relevant, and title terms are thus useful in providing focused, though incomplete, retrieval.

Because a well-written abstract highlights the most important concepts in a document, words in abstracts can be highly valuable retrieval terms; however, abstracts can vary greatly in format and quality. Author abstracts on patent front pages thus run the gamut from the highly informative to the barren. Some patent abstracts can get so tied up with the legalese of patent claims that they become nearly useless for searching. Abstracts in CA tend to be useful for searching purposes, but CAS makes those abstracts available for search on-line only in the STN version of the CA database.

The text of patent claims is especially important for infringement searching. Several databases make available the complete claims of U.S. patents covering a period exceeding the life span of U.S. patents. Representative claims for European, French, and German patents, as well as PCT applications, are available on-line. The long-awaited full claims for EP documents are expected to be a great aid in carrying out infringement searches of patents obtained through the EPO.

The full text of patent specifications is an intriguing retrieval tool. At its best, it enables the location of the tiniest detail in patent disclosures, details that can easily escape the attention of the document analysts who abstract and index patents. At its worst, it provides discussions of prior art or alternative procedures that have no relation to the invention at hand. Full-text databases such as DIALOG's or STN's U.S. files, which enable a search to be limited to portions of the patent text, can help improve the quality of full-text searches. Nevertheless, there are concepts embodied in drawings or structural diagrams that cannot be expressed by natural language search of full text. Further, there are aspects of rigorous and less-than-rigorous chemical nomenclature that present considerable challenges to the search of natural language text.

Controlled indexing can help overcome the vagaries of free text. The structures of indexing languages can differ sharply, and thus have a substantial effect on retrieval techniques. For example, the indexing language of the API databases is hierarchical. Terms have an interrelationship with broader, narrower, or related terms. When a document is indexed, it is indexed at the most specific level, and every indexing term generates all broader terms in the hierarchy, as well as selected related terms. Thus a search can be carried out as narrowly or as broadly as desired. The reactions of methane or of benzene can be sought specifically; consideration of the reactions of aliphatic hydrocarbons can be done in the confidence that all documents indexed for the reactions of methane would be retrieved. All documents on the production of hydrocarbons can be looked for with the assurance that everything indexed

for making either methane or benzene would be retrieved.

A different situation pertains to databases such as CA or the indexed CLAIMS files. In CA, there are generic and specific terms, but a broad generic term cannot normally be searched with confidence that all of the specifics that fall into the class will be retrieved. It is necessary to build up groups of homologues and synonyms in a search strategy, although the introduction of polymer class terms in the CAS Registry is a great help in carrying out broad searches of polymer information. Groups of homologues and synonyms must in general be constructed in searching the Uniterm database as well; the IFI thesaurus provides helpful listings of related terms that are useful in preparing search strategy, and the company has added to the database a number of collection terms, such as zeolites and addition polymers, that are now posted to the index whenever a specific, narrower index term is used, and that have been pack-posted to the file as an aid in imposing broad criteria on searches.

Patent classification codes are another subject-search parameter available in most patent databases. IPC codes are usually present and U.S. codes exist in a number of files; in the case of Japan Patent Information Organization (JAPIO), Japanese codes too are available. It is possible to mimic a hand search by limiting operations to references falling within one class or group of classes. Although such strategies can in some instances be justified, it is usually wiser to treat class codes as just one of the various subject parameters that make up a search strategy.

Structure Searching. Fragmentation systems have been the traditional means for indexing and searching generic and Markush chemical structures in patents. Derwent's FARMDOC-AGDOC-CHEMDOC code is such a system, as are the systems used in CLAIMS-Uniterm and -CDB and in APIPAT, but there are important differences among the systems. The Derwent system is geared only to products, not starting materials or other reactants, and thus does not include a system of roles. Fragments of a molecule are linked together to distinguish them from fragments of other substances in the same document. Uniterm has a rudimentary fragmentation system with no linking capability, so that all functional groups in all substances in the patent are thrown into the same mix. The CDB fragment system is highly detailed, and features both links to isolate the components of individual compounds and roles to denote their use. Frequently appearing chemicals that have their own descriptors are not fragmented, however, so that a search for all fluorinated alcohols must include not only the fragments for fluorine and alcohol, but also the terms for any individual fluoroalcohols contained in the indexing vocabulary. The APIPAT fragmentation system is less detailed than either Derwent's or CLAIMS-CDB's, but all indexed substances are fragmented, so that a purely generic search can safely be run for the fluoroalcohols without having to specify individual compounds. Another valuable APIPAT feature is their so-called template system of inputting the indexing for Markush formulations, which aids in the generation of multiple linked-term sets and avoids the overcoding that produces false retrieval (23).

Fragmentation systems, useful as they are, describe molecular structures incompletely. Topological indexing systems, typified by the CAS Registry, are used to identify unambiguously each of the more than 13 million substances covered in CA, and can be searched for specific complete structures as well as for substructures. With the advent of the MARPAT system in 1988 the Registry began to handle generic and Markush structures as well (24,25). In the meantime, Derwent created the WPIM Markush database to deal with both exact and inexact structures in Sections B, C, and E of CPI, and INPI created PHARMSEARCH. The Registry and MARPAT are searched by Messenger software on the STN system, whereas Questel's version of the Registry, the Derwent WPIM database, and PHARMSEARCH are searched by the DARC and Markush DARC system (26,29). ORBIT and DIALOG have gateways to WPIM; ORBIT and DIALOG also have Registry dictionary files, CHEMDEX and CHEMNAME, respectively, which enable the compounds in the CAS Registry to be searched by a combination of parameters such as name fragments, molecular formulas, and ring system identifiers. Dictionary searching may lack some of the power of the full Registry database but it is a highly useful technique in its own right, and can be used in combination with topological searching in the STN version of the Registry.

Proteins and nucleic acids present special problems for structure searching in that they contain a small group of repeating subunits but few variable groups. Sequences of polypeptides and nucleic acids resisted efforts to search them by fragmentation or topological methods until the late 1980s, when Derwent and IntelliGenetics devised GENESEQ and CAS invented a system for sequence searching within the Registry (30).

Cross-File and Multifile Techniques

Databases differ in their strengths and weaknesses, as well as in their focus. As a result, duplicate searches carried out on different databases generally produce different results. This has been demonstrated in comparative studies of retrieval results for a group of patent databases (31,32). Participants in one study (31) made an effort to develop optimal search strategies in each database tested, yet in no instance did one file produce perfect retrieval. Both investigations (31,32) found that results from the various databases complemented each other. As a result, searchers are counseled to use multiple databases whenever possible. There is no pat answer to the question of how many files to use or which files to use; however, more files mean more expenditure, and searchers must develop their own cost-benefit relationship.

It is especially valuable to be able to bring elements of one database to bear on another. This technique is known as cross-file searching. Cross-file searching had its origins in the early days of CAS Registry dictionary files on the ORBIT and DIALOG systems. Searches of those files produced lists of compound registry numbers, which had to be rekeyed for searching in the versions of the CA database that these hosts offered. ORBIT developed a technique that enabled the output of a search to be obtained directly as search terms, formatted for use in another database. ORBIT's PRINT SELECT software was followed by DIALOG's MAP software which provided an alternative method for carrying out cross-file operations. STN and Questel too have developed cross-file softwares. Each of the cross-file systems has its own special characteristics. For example, STN's SmartSELECT and Questel's MEM software enable selection of only those terms meeting criteria established by the searcher, such as only the U.S. patents in a set.

Cross-file techniques permit searchers to combine the approaches and capabilities of different databases and achieve a synergistic result (33). Thus, the APIPAT database features a vocabulary that contains many specialized terms from the petroleum refining industry and has a rudimentary chemical fragmentation code. WPI, on the other hand, has a fragmentation code that is much more precise. A set of candidate references can be developed on APIPAT by searching concepts, then passed against the WPI fragmentation code using cross-file techniques, to produce a search more precise than could be done on either database alone. An example of how references can be lost when one database focuses on just one of two retrieval parameters and a second database focuses on just the other parameter has been given (32). This type of situation can often be remedied by cross-file techniques.

Because API indexing is done from Derwent abstracts, and APIPAT and WPI references are cross-referenced to each other, the inherent ties between APIPAT and Derwent references along with experience in cross-file searching between APIPAT and WPI on the ORBIT system, have led Derwent, API, and ORBIT to undertake the physical merger of APIPAT with WPI (16). This enabled the development of search strategies by simply combining API and Derwent parameters, and confirmed the value of searches that combine the viewpoints of different databases. The desirability of a master patent database combining the features of all of the principal patent files has been expressed (31,34).

Multifile searching differs from cross-file searching in that it permits a single strategy to be brought to bear on more than one file at the same time, but the individual files are searched independently without interaction. Most on-line hosts in the 1990s permit some form of multifile searching, which is exemplified by DIALOG's OneSearch. An advantage of multifile searching is that it is possible to create and use in one step a single strategy; a disadvantage is that the single strategy may not be optimum for all of the files used. ORBIT's version of multifile searching, called PowerSearch, can introduce an element of cross-file searching to multifile situations (35,36). Different databases can be searched separately, each by whatever strategy desired, and the outputs can be merged. PowerSearch can then group references from the same or different databases that refer to the same patent or patent family. By printing the output in groups, the searcher is able to approximate an implied file merger or cross-file operation, since patents or patent families retrieved from file A by strategy A and from file B by strategy B are grouped together, and the specified features from files A and B are combined.

Term Extraction and Analysis Software. Closely allied to the software used in cross-file searching is the software that extracts terms and provides statistical analysis of their occurrence within the set being analyzed. For example, a searcher may carry out a state-of-the-art search and obtain a listing of the patentees represented, ranked by the number of patents for each. ORBIT's GET software, originally used at Pergamon Infoline, a predecessor company, was the pioneer; other systems include DIALOG's RANK, STN's SELECT, and Questel's MEMSORT. STN has also developed an offshoot of the SELECT command called SmartSELECT, which temporarily exits a file and revisits a file searched earlier to extract search terms from answer sets. Both RANK and SELECT enable the searcher to manipulate intermediate results to refine the product. Thus, if the output of some patentees is fragmented in the patentee ranking because of variations in spelling, such variations can be combined and the data rerun without incurring additional charges for the operation. Software of this type has become increasingly popular among searchers who need to analyze statistically various aspects of search

results.

Patent Databases

Derwent World Patents Index (WPI) and WPI Markush. Derwent's in-depth documentation services began in 1963 with the FARMDOC service for pharmaceuticals, followed by AGDOC for agricultural chemicals. PLASDOC covering polymers began in 1966. FARMDOC and AGDOC featured a chemical fragmentation code searched on IBM punch cards or corresponding computer tapes. A fragmentation code developed by the U.S. PTO was adopted by Derwent to handle steroid molecules. PLASDOC had its own punch card code. The code systems were severely limited because of the restrictions placed by the 960 positions on the punch card; there was considerable grouping of concepts that might better have been separated, as well as overcoding of alternatives on the same card rather than on separate cards. These limitations produced false retrieval, but such false retrieval was not a severe problem in the early days of the system, which contained modest numbers of references in the databases. Besides the punch code system, these early databases featured a manual code system for the manual searching of classified sets of documentation abstracts on search cards.

In 1970, Derwent extended its coverage to all aspects of chemistry, in the *Central Patents Index*, later renamed the *Chemical Patents Index* (CPI). The general chemical section of CPI, called CHEMDOC, was coded by a slightly modified version of the punch code system from FARMDOC and AGDOC, and manual code systems were developed for all of the nine new CPI sections. The increase in false retrieval from the growing file provided the driving force for code improvements in 1970, 1972, and 1981. Nonchemical patents were added by Derwent in 1974, expanding the total to the *World Patents Index*. Manual codes were not at first created for the nonchemical parts of the system, but were later added for electrical patents.

With the advent of on-line searching in the 1970s, the Derwent file was one of the first to go on-line. It had subject retrieval capability by the manual and punch code systems, title terms, IPC, and broad subject groupings called Derwent classes, whose primary function had been to allocate patents to appropriate segments of the Derwent system. By 1981, abstracts were added to the database, after which abstracts for the entire back-file were added. Freed from the tyranny of the IBM punch card, Derwent gradually added retrieval parameters such as inventors and multiple patentees for jointly held patents, and strengthened the capabilities of both chemical and polymer retrieval systems. A limited number of specific chemical compounds had been directly searchable since 1981. The WPIM file, which permits structural search of Markush structures by the Markush DARC system and direct searching for additional specific structures, was added for references from 1987 onward. Polymer indexing was greatly enhanced with the system introduced in 1993, which featured a greatly increased indexing vocabulary and the unique capability to link data elements on three distinct levels.

An important aspect of Derwent's treatment of bibliographic data has been a standardized method for registering data elements such as patent, priority, and application numbers. Issuing countries vary in the way they assign such numbers. The number of digits can vary, and digits for the filing or publication year may precede or follow the serial number. INPADOC standards have in many instances preserved these inconsistencies, so that searchers may be uncertain as to how to enter data in a search. Derwent has standardized the presentation of these data elements, in part because of the constraints created by computer programs used before 1992. The resulting system did in some instances dictate the elision of a letter or digit, but once understood was totally self-consistent. The Derwent format has become the standard for searching patent data across the ORBIT and Questel systems, and is an optional standard on STN. Newer computer programs have permitted Derwent to restore the elided characters since 1992.

The WPI database offers a wide range of bibliographic and subject-based access points. Indexing is deepest in the four CPI sections having special systems, ie, FARMDOC, AGDOC, CHEMDOC, and PLASDOC. Among the first three, the origins of the system in the product-oriented pharmaceutical industry have produced a system that works less well in petrochemicals, eg, where the focus is frequently on starting materials which are often not indexed in the system. But with its combination of multiple access capabilities, detailed documentation abstracts, extensive patent family information, as well as archival and document delivery capabilities, Derwent is indisputably the overall leader in providing patent information.

Chemical Abstracts and CAS Registry. CAS is especially notable for the thoroughness and high quality of its products. The CAS Registry system does a superb job of identifying any chemical that is either involved in new chemistry with hard data, or, since about 1980, specifically claimed in a patent. The MARPAT database has also led CAS to identify the perhaps nonexistent but prophetic substances covered by Markush claims in patents.

In its earliest years, the printed *Chemical Abstracts* provided lengthy abstracts that could often serve as surrogates for original documents. Derwent still does this in some instances, but CAS on the other hand strives to be a pointer to original documents. The volume of nonpatent and patent literature covered in CA dictates that complete abstracts are impractical, although CA indexing records can contain a wealth of detail not hinted at in abstracts.

Given the value of abstracts as searching tools, it is regrettable that CA's abstracts are available only in the STN version of the file. The mounting of the WPI database on the STN system provides powerful cross-file search capabilities between the fragmentation-coded references in the WPI system and the specifically registered references in CA. The WPI-CA combination also facilitates searching for information on specific patents in the CA database by individuals who may have the number of an equivalent rather than the member of a patent family made basic by CAS. Many databases on the STN system, including the USPATFULL file, have been enhanced by the addition of CAS Registry Numbers.

INPADOC and EPIDOS Register. If WPI is the preeminent multipurpose file, INPADOC is the preeminent bibliographic file. Its family information is more complete than that of any other database, although it has less intellectual family data than WPI. The legal status data, whether in a combined file as in DIALOG or a separate file as in ORBIT's LEGSTAT, is the most extensive available, and will most likely extend to additional countries soon. The subject searching capability of the database is sharply limited to only original title words (many in a variety of languages) and IPCs, but in appropriate circumstances INPADOC can still be used effectively for subject searching. For example, if an organization has recently obtained a patent in a given area of technology, and the patent publication is so new that it has not yet reached the WPI database, an INPADOC search can be quite fruitful. This is especially so with a file such as ORBIT's INPANNEW or STN's INPAMONITOR, which include just a few weeks' worth of the most recent information, and process rapidly as a result.

The EPIDOS Register available directly on the EPIDOS computer is unique in providing behind-the-scenes information, such as the expected date of grant for a pending patent application, and a record of correspondence between examiner and applicant. This file is somewhat user-unfriendly, and as a result probably has limited use; however, modifications that could increase its use are likely.

EDOC. EDOC, available on just the Questel system, provides in most instances less complete patent family information than either WPI or INPADOC, but it is unique in including Japanese C-level documents and also some Australian data not available in the other two databases. It is possible on occasion to find family information from before 1970 in this database, though such data are not provided systematically. The ability to search by the EPO-modified version of IPC is unique. Its value is tempered by the fact that output does not even include titles, much less abstracts of patents. That output can, however, be transferred by cross-file search protocols to more detailed files, such as WPI.

CLAIMS Databases. IFI Plenum offers a family of files under the CLAIMS rubric. At the heart are the three basic patent files, CLAIMS-Biblio, -Uniterm, and -CDB. The Biblio file presents front page and claims language from the patents as issued, with no added indexing. IFI does add value in a number of ways, however, which include standardized assignee names, class codes for reclassified patents, flags for reassigned and expired patents, citation counts, and searchable two-dimensional depictions of chemical structures in the claims. The Uniterm version includes indexing; it is substantial for concepts but limited for chemical structures. For subscribers, the Comprehensive Data Base has, among other things, a highly detailed structural fragmentation system and a unique system of roles applied to the indexing of polymers. The CDB system outstrips the Uniterm system by a substantial margin, and subscribers who rely on it wish that there were more subscribers so that the cost of refinements could be more easily supported. Besides the three patent files there are a number of auxiliary files. Reexamination, reassignment, and expiration data are included in the CLAIMS-Reexamination file, and the CLAIMS-Citation files on DIALOG are the only current source for searchable citation information on pre-1970 patents. The CLAIMS Compound Registry is an aid in searching the indexed files for specific compounds.

The time span covered by CLAIMS is unique: chemical patents from 1950, nonchemical patents from 1963. The bibliographic information for pre-1970 patents is unfortunately replete with errors, especially with respect to inventor names. On the other hand, IFI has done an admirable job of standardizing patent assignee names and correcting discrepancies and errors in the originals.

Full-Text Patent Databases. The LEXPAT database on the LEXIS- NEXIS system, the first commercially available full-text patent file, receives its greatest use from patent attorneys and has been relatively unused by other patent information specialists. This may be attributed to search software that is quite different from the type familiar to information specialists, no matter what their preferred host system. This situation has changed with DIALOG's release of the PATFULL files followed by STN's USPATFULL, both searchable by familiar Boolean techniques and featuring greater

capability for searching selected portions of patent specifications. Full patent specifications generally contain much information that is not relevant to the invention itself, and searches of full-text files, unless they are carefully framed, can produce prodigious amounts of unwanted and irrelevant answers. On the other hand, full-text searches are unique in their ability to locate the most minute passing disclosure, the type of information that can be utterly inaccessible by any other means. DIALOG's PATFULL has been enhanced by some controlled data elements from CLAIMS. STN's USPATFULL has also been enhanced by CAS indexing of the U.S. patent or its equivalent. In addition, USPATFULL contains a thesaurus of the U.S. Manual of Classification. Special file enhancements of this sort are likely to proliferate as a result of competition among database producers and hosts. An important feature of LEXPAT is that post-issuance changes such as reassignments and corrections are incorporated directly to the file; there is no need to look for them in a second place.

Other Individual Country Databases and Auxiliary Files. The USPatents files on ORBIT, supplied by Derwent, are similar in their contents to the CLAIMS-Bibliographic files, including all the front page information and the full claims language. These files do not include the two-dimensional structures provided by IFI, nor do they have IFI's standardization of assignee names. Citation searching is available at a cost considerably less than that for the portion of the CLAIMS-Citation file that covers the same period.

Several auxiliary files complement the various basic U.S. patent databases. USCLASS on ORBIT coordinates updated patent classes and patent numbers going back to the beginning of the U.S. patent system. CLAIMS Reference on DIALOG (File 124) and IFIREF on STN include the text of the U.S. classification system, along with the IFI indexing vocabulary; CLAIMS-CLASS on ORBIT has just the classification text, not the indexing vocabulary. CLAIMS-Registry shows the fragmentation of compounds specifically indexed in the IFI system, and a generic search of this file produces a search list of specific compounds to augment a generic fragment search. The Patent Status File from Derwent's Rapid Patent subsidiary is available in print and on the ORBIT system, and collects a wide range of post-issuance events in the life of U.S. patents, including corrections and reassignments. LITALERT, from the same source, provides information on U.S. patent litigation.

Files covering the patent output of individual patent offices have proliferated in the 1990s. They include EPAT and PATOSEP for the EPO, PCTPAT and PATOSWO for the PCT, PATDPA and PATOSDE for Germany, PATDD for the former GDR, FPAT for France, as well as ITALPAT, JAPIO, and Chinese Patent Abstracts. Some of these have unique features. FPAT goes on-line on the day of patent publication, providing the best service in timeliness. PATDPA and FPAT contain images from patents. However, ITALPAT, probably the least informative of all patent databases, includes only application numbers, assignees, inventors, and titles, without even patent numbers.

The Chinese patent file with its abstracts can be used to supplement WPI, which at this writing has only titles for Chinese patents. JAPIO provides abstracts based in particular on patent claims, and can help to clarify uncertainties with Japanese abstracts from Derwent and or CA. PATOLIS, in Japanese, is a unique source of Japanese legal status information (37). EPIDOS staff carry out PATOLIS searches on request; for those with sufficient need to search the PATOLIS database, software is available to enable those who cannot read Japanese to extract key data.

Other Databases with Patent Information. The APIPAT database has been discussed, as have the unique capabilities of the merged WPI-APIPAT file on ORBIT. Many other databases contain substantial amounts of patent information, notable among them are Derwent's *Biotechnology Abstracts*, the TULSA database (petroleum exploration and production), several specialized pharmaceutical files, PAPERCHEM, and METADEX. A very complete listing of databases containing patent information circa 1990 has been prepared (38).

Archiving and Document Delivery

Many organizations have traditionally maintained in-house collections of patent specifications in areas of interest. Microforms and, more recently, CD-ROMs have taken the place of paper copy as the volume of the patent literature grows. CD-ROMs can each hold the images for approximately 1000 U.S. or European patent documents, and high quality copies can be produced with laser printers. CD-ROMs are available soon after patent publication, and represent a significant advance over earlier patent copy delivery media, but they too may be supplanted in the future. MicroPatent and Derwent are suppliers of CD-ROMs based on U.S. patents, and serve in addition as agents for the EPO's CD-ROM products. EPIDOS produces the ESPACE series of CD-ROMs for the EPO, PCT, and several additional European countries. The Japanese patent office produces Japanese language CD-ROMs, which have completely replaced paper copies as Japan's publication medium.

Most users of patent information, occasionally even those that have extensive patent copy collections, must order copies from the outside, and national patent offices such as the U.S. PTO have been important suppliers of copies, typically for their own country and other countries as well. In the United States, there are more than 70 patent depository libraries that serve as regional patent information resources. In addition, a number of other organizations supply patent copies, including Derwent, CAS, DIALOG, Rapid Patent, The Library Connection, the British Library, EPIDOS, and others. Electronic means of document delivery are becoming more common; telefacsimile (FAX) are used by many suppliers, and satellite transmission has been discussed by some potential suppliers. The Internet has been used for the transmission of some patent information. As of the middle of 1995, a rapidly increasing number of patent-related Internet activities was beginning to appear. It seems likely that some of these activities will develop into important information resources in the future.

Other Technological Initiatives

Strategies developed using Boolean logic have been central to the searching of on-line databases, but during the early 1990s it has been suggested that effective searches could be carried out with natural language input by using systems based on artificial intelligence. Two of these in particular are related to patent searching. DIALOG's TARGET software is aimed not particularly at patents, but at any database containing substantial amounts of text, especially full-text files. Full text is desirable because the system is based in part on term-frequency counts. A second system, presumably of similar type, is Patent Analyzer, offered jointly by Derwent and Electronic Data Systems. Both are so new that there is insufficient information for a meaningful evaluation of their effectiveness. If successful, systems of this sort could have a significant effect on the way in which patent searching is done in the future. It would appear, however, that such systems might be less appropriate for searching the chemical arts than for those functions in which the recognition of chemical structures is not required.

CASLINK, a software feature from STN, carries out searches of several structure files, including Registry, MARPAT, and MARPAT Previews, collects the results, runs these against CA bibliographic databases, and identifies and eliminates duplicate records (39). More and more systems are being developed to simplify mechanical operations required for a searcher, and more capability is being developed to identify and, if desired, eliminate duplicate records. The ORBIT PowerSearch software is probably the most effective at recognizing patent family relationships, but competition among the on-line hosts and database producers will lead to continued improvement of these capabilities.

BIBLIOGRAPHY

"Patent (Literature)" in *ECT* 1st ed., Vol. 9, pp. 890–897; in *ECT* 2nd ed., Vol. 14, pp. 583–635, by E. S. Turner, Bell Telephone Laboratories, Inc.; in *ECT* 3rd ed., Vol. 16, pp. 889–945, by J. W. Lotz, IFI Plenum Data Co.

1. *Manual of Patent Examining Procedure*, Section 901.04 [Revision 14], U.S. Department of Commerce, Patent and Trademark Office, Washington, D.C.
2. Standard Code for Identification of Different Kinds of Patent Documents, "Patent Cooperation Treaty," Annex C.
3. WO 95 00525.
4. WIPO, *International Patent Classification*. 5th ed., Carl Heymans Verlag KG, Munich, Germany, 1989.
5. D. T. Dickens, *World Pat. Inf.* **16**, 28–32 (1994).
6. *Manual of Classification*, U.S. Department of Commerce, Patent and Trademark Office, Washington, D.C.; revised regularly.
7. *Patent Classification Definitions*, U.S. Department of Commerce, Patent and Trademark Office, Washington, D.C.; revised regularly.

8. S. M. Kaback, in H. F. Mark and co-workers, eds., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 10, John Wiley & Sons, Inc., New York, 1987, pp. 787–801; S. M. Kaback, *J. Chem. Inf. Comput. Sci.* **31**, 439–443 (1991).

9. J. A. Briggs, E. A. Ferns, and K. E. Shenton, *J. Chem. Inf. Comput. Sci.* **31**, 454–458 (1991); S. M. Kaback, *World Pat. Inf.* **16**, 60–61 (1994).

10. E. S. Simmons, *J. Chem. Inf. Comput. Sci.* **24**, 10–15 (1984).

11. M. Z. Balent and J. M. Emberger, *J. Chem. Inf. Comput. Sci.* **15**, 100–104 (1975).

12. N. Lambert, *Database*, **10**(6) 46–59 (1987).

13. M. P. O'Hara and C. Pagis, *J. Chem. Inf. Comput. Sci.* **31**, 59–63 (1991).

14. E. H. Brenner, *World Pat. Inf.* **9**, 27–33 (1987).

15. A. Girard, K. Landsberg, and S. M. Kaback, *Database*, **1**(2) 46–67 (1978).

16. S. M. Kaback, *Database*, **13**(2) 22–26 (1990).

17. E. S. Simmons, *J. Chem. Inf. Comput. Sci.* **25**, 379–386 (1985).

18. E. S. Simmons, *Database*, **18**(3) 28–37 (1995).

19. M. P. Carpenter, F. Narin, and P. Woolf, *World Pat. Inf.* **3**, 160–163 (1981).

20. M. E. Mogee and R. G. Kolar, *World Pat. Inf.* **14**, 237–244 (1992).

21. S. M. Kaback, *World Pat. Inf.* **6**, 80–81 (1984); E. S. Simmons and S. M. Kaback, *World Pat. Inf.* **10**, 204–206 (1988).

22. E. S. Simmons and N. Lambert, in H. Collier, ed., *Recent Advances in Chemical Information*, CRC Press, Inc., Boca Raton, Fla., 1992, pp. 33–78.

23. S. M. Kaback, *World Pat. Inf.* **16**(2) 1994.

24. W. Fisanick, *J. Chem. Inf. Comput. Sci.* **30**, 145–154 (1990).

25. T. Ebe, K. H. Sanderson, and P. S. Wilson, *J. Chem. Inf. Comput. Sci.* **31**, 31–36 (1991).

26. K. A. Cloutier, *J. Chem. Inf. Comput. Sci.* **31**, 40–44 (1991).

27. E. S. Simmons, *J. Chem. Inf. Comput. Sci.* **31**, 45–53 (1991).

28. N. R. Schmuff, *J. Chem. Inf. Comput. Sci.* **31**, 53–59 (1991).

29. J. M. Barnard, *J. Chem. Inf. Comput. Sci.* **31**, 64–68 (1991).

30. A. H. Berks, *Trends Biotech.* **12**, 352–364 (1994).

31. S. M. Kaback, *Online*, **7**(4) 22–31 (1983).

32. B.-R. Weckend, *World Pat. Inf.* **9**, 140–146 (1987).

33. S. M. Kaback, *Database*, **10**(5) 17–30 (1987).

34. M. A. Lobeck, *World Pat. Inf.* **16**, 14–27 (1994).

35. K. A. Cloutier, *Online*, **18**(3) 85–92 (1994).

36. N. Lambert, *Searcher*, **2**(5) 1994.

37. G. Vacek, *World Pat. Inf.* **16**, 41–45 (1994).

38. J. F. Sibley, *Online Patents, Trade Marks, and Service Marks Databases*, ASLIB, London, 1991.

39. R. E. Buntrock, *Database*, **17**(3) 116–121 (1994).

General References

Manual of Patent Examining Procedure. U.S. Department of Commerce, Patent and Trademark Office, Washington, D.C.; revised periodically.

M. A. Epstein, *Modern Intellectual Property*, 3rd ed., Aspen Law and Business, Englewood Cliffs, N.J., 1995.

B. A. Amernick, *Patent Law for the Nonlawyer: A Guide for the Engineer, Technologist, and Manager*, 2nd ed., Van Nostrand Reinhold, Co., Inc., New York, 1991.

A. Reverdin and F. Schlaepfer, eds., *Katzarov's Manual on Industrial Property All Over the World*, Katzarov SA, Geneva, 9th ed., 1981; 11th rev., 1991.

J. Maynard and H. Peters, *Understanding Chemical Patents*, 2nd ed., American Chemical Society, Washington, D.C., 1991.

B. M. Rimmer, *International Guide to Official Industrial Property Publications*, 3rd ed., revised and updated by S. Van Dulken, British Library, London, 1992.

P. D. Rosenberg, *Patent Law Basics*, Clark Boardman Callaghan, Deerfield, Ill., 1992; revised annually.

E. S. Simmons, in C. J. Armstrong and J. A. Large, eds., *Manual of Online Search Strategies*, 2nd ed., Ashgate, Aldershot, U.K., 1992, pp. 51–127.

S. Van Dulken, ed., *Introduction to Patents Information*, 2nd ed., British Library, London, 1992.

W. Warr and C. Suhr, *Chemical Information Management*, VCH Publishers, New York, 1992.

WIPO, "Glossary of Terms Concerning Industrial Property Information and Documentation," *World Pat. Inf.* **15**, 21–39 (1993).

Edlyn S. Simmons
Hoechst Marion Roussel, Inc.

StuartM. Kaback
Exxon Research and Engineering Co.

PBI FIBERS.

See HIGH PERFORMANCE FIBERS.

PCBs.

See CHLOROCARBONS AND CHLOROHYDROCARBONS, TOXIC AROMATICS.

PEAT.

See LIGNITE AND BROWN COAL.

PECTIC SUBSTANCES.

See GUMS.

PELLETING AND BRIQUETTING.

See SIZE ENLARGEMENT.

PENICILLINS.

See ANTIBIOTICS, β LACTAMS PENICILLINS AND OTHERS.

PENTAERYTHRITOL.

See ALCOHOL, POLYHYDRIC.

PENTANES.

See HYDROCARBONS.

PEPTIDE ANTIBIOTICS.

See ANTIBIOTICS, PEPTIDES.

PEPTIDES.

See BIOPOLYMERS; HORMONES; PROTEIN ENGINEERING; PROTEINS.

PERCHLORIC ACID AND PERCHLORATES

[7601-90-3], HClO_4 , one of the strongest of the mineral acids. The perchlorates are more stable than the other chlorine oxyanions, ie, chlorates, ClO_3^- ; chlorites, ClO_2^- ; or hypochlorites, OCl^- (3) (see CHLORINE OXYGEN ACIDS AND SALTS). Essentially, all of the commercial perchlorate compounds are prepared either directly or indirectly by electrochemical oxidation of chlorine compounds (4–8) (see ALKALI AND CHLORINE PRODUCTS; ELECTROCHEMICAL PROCESSING). The perchlorates of practically all the electropositive metals are known, except for a few cations having low charges.

The most outstanding property of the perchlorates is their oxidizing ability. On heating, these compounds decompose into chlorine, chlorides, and oxygen gas. Aqueous perchlorate solutions exhibit little or no oxidizing power when dilute or cold. However, hot concentrated perchloric acid is a powerful oxidizer and whenever it contacts oxidizable matter extreme caution is required. The acidified concentrated solutions of perchlorate salts must also be handled with caution. Ammonium perchlorate [7790-98-9] (AP) is one of the most important perchlorates owing to its high (54.5% O_2 content and the

absence of residue on decomposition. These properties, along with a long shelf life, make it a useful rocket propellant (see EXPLOSIVES AND PROPELLANTS) (9). AP is a true explosive as demonstrated by the explosion at the PEPCON plant at Henderson, Nevada, in 1988 (10).

Following early (1890s) work (11), France, Germany, Switzerland, and the United States began to produce perchlorates for use as propellants and explosives. Whereas total world production of perchlorates did not exceed 1800 t yr until 1940, it increased dramatically during World War II to about 18,000 t yr in order to supply the rocket and missile industries. Actual perchlorate production is difficult to determine in any given year, because AP is classified as a strategic material. Future production is expected to depend mostly on space programs.

Properties

Chlorine Heptoxide. The anhydride of perchloric acid is chlorine heptoxide [10294-48-1], Cl₂O₇, also known as dichlorine heptoxide. It is obtained as a colorless oily liquid by dehydration of perchloric acid using a strong dehydrating agent such as phosphorus pentoxide, P₂O₅ (12,13):



The Cl₂O₇ decomposes spontaneously on standing for a few days. The acid dehydration reaction requires a day for completion at 10°C and explosions can occur. Upon ozonation of chlorine or gaseous ClO₂ at 30°C, Cl₂O₇ is formed (13).

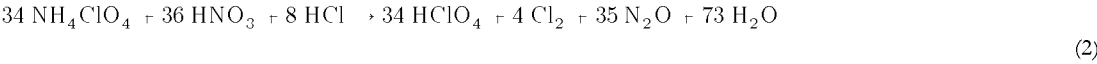
Chlorine heptoxide is more stable than either chlorine monoxide or chlorine dioxide; however, the Cl₂O₇ detonates when heated or subjected to shock. It melts at 91.5°C, boils at 80°C, has a molecular weight of 182.914, a heat of vaporization of 34.7 kJ mol (8.29 kcal mol), and, at 0°C, a vapor pressure of 3.2 kPa (23.7 mm Hg) and a density of 1.86 g mL (14,15). The infrared spectrum is consistent with the structure O₃ClOClO₃ (16). Cl₂O₇ decomposes to chlorine and oxygen at low (0.2–10.7 kPa (1.5–80 mm Hg)) pressures and in a temperature range of 100–120°C (17). It is soluble in benzene, slowly attacking the solvent with water to form perchloric acid; it also reacts with iodine to form iodine pentoxide and explodes on contact with a flame or by percussion. Reaction with olefins yields the impact-sensitive alkyl perchlorates (18).

Perchloric Acid. Pure anhydrous perchloric acid, HClO₄, is quite unstable. In aqueous solution, however, HClO₄ is a familiar and useful reagent. The acid is the strongest simple acid and the perchlorate ion the least polarizable negative ion known. Perchloric acid is commonly obtained as an aqueous solution, although the pure anhydrous compound can be prepared by vacuum distillation as a colorless liquid, which freezes at 112°C and boils at 16°C at 2.4 kPa (18 mm Hg) without decomposition. The pure acid cannot be distilled at ordinary pressures and explodes at 90°C after standing at room temperature for 10–30 days. The aqueous solution can be concentrated by boiling at 101 kPa (1 atm) at 203°C, at which point an azeotropic solution is attained which contains 72.4% HClO₄. For purification by distillation, reduced pressure is needed below 200 mm to avoid partial decomposition to chlorine, chlorine oxides, and oxygen (19–24).

A number of hydrates of perchloric acid, HClO₄·*n*H₂O, where *n* = 1, 2, 2.5, 3, and 3.5, are known. These are commonly referred to as the hydronium or oxonium perchlorates, H₃O⁺ClO₄[−], because of the analogy between the x-ray patterns of these species and ammonium perchlorate. Commercial 72% perchloric acid contains only slightly more water than the dihydrate (25). When cold or dilute, perchloric acid is a weak oxidizing agent. When hot and concentrated, however, its oxidizing power increases to the point that it can act explosively in the presence of a reducing agent (26).

The combination of oxidizing effect, acidic strength, and high solubility of salts makes perchloric acid a valuable analytical reagent. It is often employed in studies where the absence of complex ions must be ensured. The value of wet ashing techniques, in which perchloric acid is used to destroy organics prior to elemental analysis for the determination of trace metals in organics, has been well established (see TRACE AND RESIDUE ANALYSIS).

Perchloric acid can be prepared by the treatment of perchlorates with sulfuric acid followed by distillation. A modification of the procedure (21) involves the reaction of ammonium perchlorate with nitric and hydrochloric acids, and then concentration at 198–200°C to eliminate the unreacted acids by vacuum distillation:



The electrolytic oxidation of chlorate to perchloric acid is also feasible (27). Perchlorates are commonly prepared by electrolytic oxidation of chlorates:



Electrochemical perchloric acid formation begins at 2.4 volts and reaches a maximum at 2.8–2.9 volts; lowering the temperature to 20°C can accelerate the process significantly (28).

When solid potassium chlorate is carefully heated, it can be transformed into perchlorate thermally:



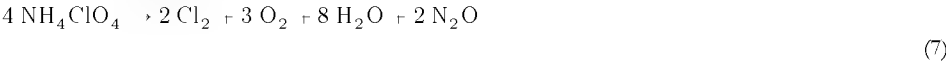
Ammonium Perchlorate. Heats of formation for the metal perchlorates are nearly the same as those for the corresponding chlorides, so that the reaction



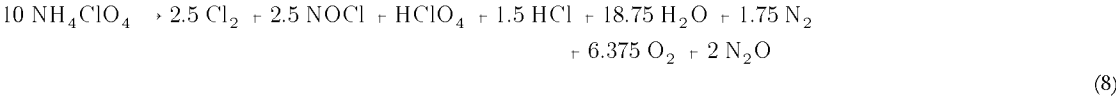
takes place with little net energy change. It is for this reason that the perchlorates, especially those of the light metals and ammonium ion, are favored as solid oxidizers for rocket propellants.

Ammonium perchlorate is a colorless, crystalline compound having a density of 1.95 g mL and a molecular weight of 117.5. It is prepared by a double displacement reaction between sodium perchlorate and ammonium chloride, and is crystallized from water as the anhydrous salt.

Because of the use of ammonium perchlorate as a solid oxidizer for rocket propellants, the thermal decomposition has been much studied (29–32). Three separate activation energies have been observed for AP decompositions: an activation energy of 123.8 kJ mol (29.6 kcal mol) is found below 240°C; of 79.1 kJ mol (18.9 kcal mol) above 240°C; and finally, of 307.1 kJ mol (73.4 kcal mol) between 400–440°C (33,34). Below 300°C, the equation



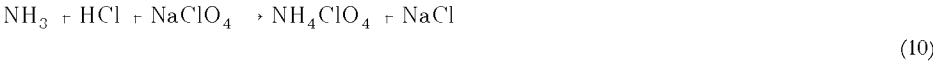
represents the primary products. Above 300°C, the proportion of nitric oxide increases; and above 350°C, the gas analysis suggests the following (32):



In commercial manufacture of ammonium perchlorate, sodium perchlorate can be the starting material. The ammonium ion can be contributed by such materials as ammonium chloride, sulfate, and nitrate, eg, the metathetical reaction of sodium perchlorate and ammonium chloride:



In a modification of equation 9 (35),



A newer approach developed for producing commercial quantities of high purity AP (8,36) involves the electrolytic conversion of chloric acid [7790-93-4] to perchloric acid, which is neutralized by using ammonia gas:



The ammonium perchlorate solution is spray-dried to the desired crystal size at air temperatures below 150°C and crystal temperatures of about 110°C. This procedure provides a pure product having a controlled grain size. Prior mechanical and thermal treatment affects the isothermal decomposition of AP at 215–235°C (37).

Alkali Metal Perchlorates. The anhydrous salts of the Group 1 (IA) or alkali metal perchlorates are isomorphous with one another as well as with ammonium perchlorate. Crystal structures have been determined by optical and x-ray methods (38). With the exception of lithium perchlorate, the compounds all exhibit dimorphism when undergoing transitions from rhombic to cubic forms at characteristic temperatures (33,34). Potassium perchlorate [7778-74-7], KClO₄, the first such compound discovered, is used in pyrotechnics (qv) and has the highest percentage of oxygen (60.1%).

The alkali metal perchlorates are either white or colorless, and have increasing solubility in water in the order of Na > Li > NH₄ > K > Rb > Cs. The high solubility of sodium perchlorate, NaClO₄, makes this material useful as an intermediate for production of all other perchlorates by double metathesis reactions and controlled crystallization.

Group 11 (IB) Perchlorates. Copper and silver perchlorates have been studied quite extensively. Copper(I) perchlorate [17031-33-3], CuClO₄, and copper(II) perchlorate [13770-18-8], Cu(ClO₄)₂, form a number of complexes with ammonia, pyridine, and organic derivatives of these compounds. The copper perchlorate is an effective burn-rate accelerator for solid propellants (39).

The silver perchlorate [7783-93-9] salt, AgClO₄, is deliquescent and forms a light-sensitive monohydrate that can be dehydrated at 43°C and is soluble in a variety of organic solvents. Explosions of silver perchlorate have been reported (40–42). Gold forms organic perchlorate [42774-61-8] complexes as well as complexes with silver, eg, (C₆H₅)₃AgAu(C₆F₅)₂ClO₄.

Alkaline-Earth Perchlorates. Anhydrous alkaline-earth metal perchlorates can be prepared by heating ammonium perchlorate in the presence of the corresponding oxides or carbonates (43). The hydrates can be prepared by treatment of the metal oxides or various salts with aqueous perchloric acid (44). The alkaline-earth perchlorates are unusually soluble in organic solvents (44). The basic salts M(OH)ClO₄, where M is Mg, Ca, or Ba, have also been prepared and characterized (45). Beryllium perchlorate [39455-86-2], Be₄O(ClO₄)₄, has been prepared by allowing the BeCl₂ salt and HClO₄ to react or by heating BeCl₂ and HClO₄·H₂O at 60°C (46). The dihydrate forms.

Group 12 (IIB) Perchlorates. The zinc perchlorate [13637-61-1], cadmium perchlorate [13760-37-7], mercury(I) perchlorate [13932-02-0], and mercury(II) perchlorate [7616-83-3] all exist. Cell potential measurements show that zinc and cadmium perchlorates are completely dissociated in concentrations up to 0.1 molar in aqueous solutions (47–49). Mercurous perchlorate forms a tetrahydrate that can be readily converted to the dihydrate on heating to above 36°C (50).

Group 13 (IIIA) Perchlorates. Perchlorate compounds of boron and aluminum are known. Boron perchlorates occur as double salts with alkali metal perchlorates, eg, cesium boron tetraperchlorate [33152-95-3], Cs(B(ClO₄)₄) (51). Aluminum perchlorate [14452-95-3], Al(ClO₄)₃, forms a series of hydrates having 3, 6, 9, or 15 moles of water per mole of compound. The anhydrous salt is prepared from the trihydrate by drying under reduced pressure at 145–155°C over P₂O₅ (52).

Group 3 (IIIB) and Inner Transition-Metal Perchlorates. The rare-earth metal perchlorates of yttrium and lanthanum have been reported (53). Tetravalent cerium perchlorate [14338-93-3], Ce(ClO₄)₄, and uranium perchlorate have also been identified (54).

Group 14 (IVA) Perchlorates. Perchlorates containing organic carbon have been reported, as have diazonium perchlorates, oxonium perchlorates, and the perchlorate esters (55–57). Extreme caution must be used in working with organic perchlorates; many decompose violently when heated, contacted with other reagents, or subjected to mechanical shock. The diazonium perchlorate of *p*-phenylenediamine, ClO₄N₂(C₆H₄)N₂ClO₄, was reported in 1910 to be the most explosive substance known (58).

Group 4 (IVB) Perchlorates. Titanium tetraperchlorate [13498-15-2] sublimes at 70°C, decomposes on aging in a vacuum, and explodes when heated at atmospheric pressure to 130°C (59).

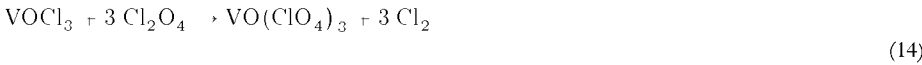
Group 15 (VA) Perchlorates. Nitrogen perchlorates have been used as oxidizers in rocket propellants. Hydrazine perchlorate [13762-80-6], NH₂NH₃ClO₄, and hydrazine diperchlorate, ClO₄NH₃NH₃ClO₄, have been investigated as oxidizers for propellant systems (60). Anhydrous salts can be recrystallized from ethanol, where the monopерchlorate adduct melts at 137–138°C and begins to decompose at 145°C. Deflagration results with rapid heating; violent detonation occurs with mechanical impact, shock, or friction.

Nitronium perchlorate, NO₂ClO₄, also called nityrl or nitroxyl perchlorate, is prepared by reaction of dinitrogen pentoxide and anhydrous perchloric acid. This nitrogen-containing perchlorate reacts vigorously with many organic compounds, explosively with some (61). Nitrosyl perchlorate [15605-28-4], NOClO₄, is found from Raman spectroscopy to be composed of NO⁺ and ClO₄⁻ ions in an orthorhombic crystal structure of density 2.169 g/mL (62–64). Reaction with water produces nitrogen oxides; reaction with methanol, nitromethane. The NOClO₄ is made by passing nitric oxide and nitrogen dioxide in 72% perchloric acid. The anhydrous salt is obtained by partially drying the hydrate over P₂O₅ in an atmosphere of nitrogen oxides followed by desiccation in vacuum. Decomposition occurs at 100°C according to the following reaction:



Phosphonium perchlorate, P(OH)₄ClO₄, can be formed as a crystalline product that melts at 46–47°C. This compound is obtained by mixing phosphoric and perchloric acids (65).

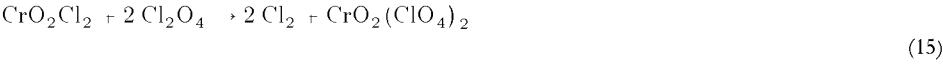
Group 5 (VB) Perchlorates. Vanadyl perchlorate [67632-69-3], VO(ClO₄)₃, has been prepared in the cold (–45 to 20°C) by the following reaction (66):



Group 16 (VIA) Perchlorates. A perchlorate compound perchloryl sulfate [43059-05-8], SO₄(ClO₄)₂ was produced by the low temperature

electrolysis of a 12-*N* H₂SO₄ and 3-*N* HClO₄ solution. This compound is a strong oxidizer; reaction with toluene, acetone, benzene, or alcohol at room temperature produces an exothermic and explosive reaction. The SO₄(ClO₄)₂ is soluble in Freon and CCl₄ without reaction (67).

Group 6 (VIB) Perchlorates. Both divalent and trivalent chromium perchlorate compounds [13931-95-8; 13527-21-9] have been reported. Anhydrous chromyl perchlorate [60499-74-3] has been prepared in the cold:



Chromyl perchlorate has been suggested for a gas-generating system operating at -45°C (66).

Group 17 (VIIA) Perchlorates. Fluorine perchlorate [37366-48-6], FClO₄, is formed by action of elemental fluorine and 60–70% aqueous perchloric acid solution (68). The compound is normally a gas. It melts at -167.5°C and boils at -15.9°C. It is extremely reactive and explosive in all states.

The perchloryl fluoride [7616-94-6], FClO₃, the acyl fluoride of perchloric acid, is a stable compound. Normally a gas having a melting point of -147.7°C and a boiling point of -46.7°C, it can be prepared by electrolysis of a saturated solution of sodium perchlorate in anhydrous hydrofluoric acid. Some of its uses are as an effective fluorinating agent, as an oxidant in rocket fuels, and as a gaseous dielectric for transformers (69).

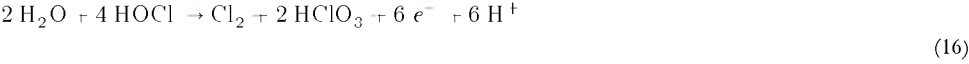
Other Transition Element Perchlorates. Both divalent and trivalent manganese perchlorate compounds [13770-16-6; 13498-03-8] are known. Perchlorates of Fe, Co, Ni, Rh, and Pd have been produced as colored crystals (70–72).

The perchlorate ion, ClO₄⁻, is considered to be noncoordinating in the presence of water. When water is rigorously excluded, anhydrous complexes such as Ni(CH₃CN)_{*n*}(ClO₄)₂, where *n* is 2, 4, or 6, can be prepared. Perchlorate complexes of Ni, Co, Cu, and Sn have been reported. In each case, however, an organic group such as CH₃CN, CH₃, or pyridyl is involved (73–76).

Manufacture

Perchloric Acid. Several techniques have been employed in the manufacture of perchloric acid, including thermal decomposition of chloric acid (77), anodic oxidation of chloric acid (8), irradiation of chlorine dioxide solutions (78), electrolysis of hydrochloric acid (79), oxidation of hypochlorites by ozone (qv) (80), ion exchange (qv), and electrodialysis of perchlorate salts (81). Perchloric acid is commercially manufactured by reacting saturated solution of sodium perchlorate with hydrochloric acid (9,82). Precipitated sodium chloride is separated from the dilute solution (32% by weight HClO₄) by filtration, and the solution is concentrated to 70% by weight via vacuum distillation. Another commercial manufacturing process involves the anodic oxidation of gaseous chlorine dissolved to about 3 g/L in 40% by weight HClO₄ at -5°C (83). The electrolysis is carried out in a filter-press-type diaphragm-separated horizontal electrolyzer. Platinum foil titanium anodes and silver cathodes are operated at 2.5–5.0 kA/m² (see METAL ANODES). The cell operates at 4.4 V at a current efficiency of 60%. The power required to transfer seven electrons at these conditions is 9600 kWh/t of perchloric acid. The platinum coating dissolves at the anode and redeposits at the cathode at an estimated rate of about 0.025 g/t of 70% perchloric acid. The high purity product manufactured by this process further allows unusual perchlorates to be prepared by direct conversion with perchloric acid.

Highly pure perchloric acid can also be produced by a patented electrochemical process in which 22% by weight hypochlorous acid is oxidized to chloric acid in a membrane-separated electrolyzer, and then additionally oxidized to perchloric acid (8,84). The desired electrochemical oxidation takes place in two stages:



The anode and cathode chambers are separated by a cation-permeable fluoropolymer-based membrane (see MEMBRANE TECHNOLOGY). Platinum-electroplated high surface area electrodes sold under the trade name of TySAR (Olin) (85,86) were used as the anode; the cathode was formed from a two-layer Hastelloy (Cabot Corp.) C-22-mesh structure having a fine outer 60-mesh structure supported on a coarse inner mesh layer welded to a backplate. The cell voltage was 3.3 V at 8 kA/m², resulting in a 40% current efficiency. The steady-state perchloric acid concentration was about 21% by weight.

Perchlorates. Historically, perchlorates have been produced by a three-step process: (1) electrochemical production of sodium chlorate; (2) electrochemical oxidation of sodium chlorate to sodium perchlorate; and (3) metathesis of sodium perchlorate to other metal perchlorates. The advent of commercially produced pure perchloric acid directly from hypochlorous acid means that several metal perchlorates can be prepared by the reaction of perchloric acid and a corresponding metal oxide, hydroxide, or carbonate.

Sodium Perchlorate. The electrochemical oxidation of sodium chlorate is carried out at the anode in an undivided cell according to the following reaction:



The standard potential for the anodic reaction is 1.19 V, close to that of 1.228 V for water oxidation. In order to minimize the oxygen production from water oxidation, the cell is operated at a high potential that requires either platinum-coated or lead dioxide anodes. Various mechanisms have been proposed for the formation of perchlorates at the anode, including the discharge of chlorate ion to chlorate radical (87–89), the formation of active oxygen and subsequent formation of perchlorate (90), and the mass-transfer-controlled reaction of chlorate with adsorbed oxygen at the anode (91–93). Sodium dichromate is added to the electrolyte in platinum anode cells to inhibit the reduction of perchlorates at the cathode. Sodium fluoride is used in the lead dioxide anode cells to improve current efficiency.

The kinds of cathodes used in industrial electrolyzers are iron, steel, or bronze. The cell tank is always constructed as the cathode and is always negatively charged to prevent cathodic corrosion. In all processes, heat must be removed from the electrolyte either by cooling the internal cell or by circulating through an external heat exchanger. Table 1 gives the operating data for a typical perchlorate cell.

Table 1. Sodium Perchlorate Cell Operating Information^a 92–101.

current, A	500–5000
current density, kA/m ²	1.5–5.2
cell potential, V	4.8–6.8
anode	PbO ₂ graphite, platinum, or Pt on copper
cathode	bronze, stainless steel (316), or iron
anode–cathode spacing, cm	0.2–3
current efficiency, %	90–97 ^b ; 85 ^c
temperature, °C	30–60

pH	6–10
sodium dichromate concentration, g L	0–5
electrolyte concentration, g L	
initial	
NaClO ₃	100–700
NaClO ₄	0–700
end	
NaClO ₃	3–100
NaClO ₄	500–1100
energy consumption, kWh kg of NaClO ₄	2.45–3.0
platinum consumption, g t of NaClO ₄	2–7
operation mode	batch or continuous
^a Refs.	
^b Pt anodes.	
^c PbO ₂ anodes.	

The electrolyte feed to the cells is pretreated to remove impurities, and or additives are added to the feed to improve cell performance (94). The cell liquor leaving the cell is evaporated, crystallized, and centrifuged to remove solid sodium perchlorate. The clarified cell liquor can undergo reaction in a double metathesis reactor to produce NH₄ClO₄, KClO₄, or other desired perchlorates.

Producers have developed specific cell configurations to optimize electricity consumption, cell capital, and operating costs. Pacific Engineering Corp., Kerr-McGee Chemical Corp., Chedde Pechiney, Cardox Corp., Electrochemie Turgi, American Potash and Chemical, and I. G. Farbenindustrie each has a unique cell design.

Ammonium Perchlorate. The commercial AP product is manufactured by the double-exchange reaction of sodium perchlorate and ammonium chloride (102,103).



Ammonia, hydrochloric acid, and sodium perchlorate are mixed and the reaction mixture crystallized in a vacuum-cooled crystallizer. Ammonium perchlorate crystals are centrifuged, reslurried, recentrifuged, and then dried and blended for shipment. Mother liquor is evaporated to precipitate sodium chloride and the depleted mother liquor is recycled to the reactor. The AP product made by this method is 99% pure and meets the specifications for propellant-grade ammonium perchlorate. The impurities are ammonium chloride, sodium perchlorate, ammonium chlorate, and water insolubles.

Extremely high purity ammonium perchlorate can be made by the direct reaction of ammonia and pure perchloric acid solution (8,36):



The reaction mixture can either be crystallized, centrifuged, and dried, or spray-dried and cyclone-separated to produce a fine crystalline powder having a particle size of 50 μm. Metal analysis of the AP produced by this method is reported to be less than 0.02 μg/g.

Shipping and Handling

Perchloric acid and perchlorates are classified as strong oxidizers and emit toxic fumes when decomposed; contact with combustible, flammable, or reducing materials must be avoided. Perchloric acid and perchlorates must be shipped in accordance with the U.S. Department of Transportation hazardous material regulations (104). The maximum shippable quantity, type of packaging, allowable carriers, and other requirements are specified in these regulations. Some perchlorates may not be shipped by a public carrier, passenger-carrying aircraft, or railroad. Handling these compounds requires the procedures and safety precautions specified by the product supplier. Perchlorates contain a self-sustaining source of oxygen, thus fires involving perchlorates must be extinguished with water. A class of more hazardous compounds is formed by mixing inorganic perchlorates with finely divided metals, sulfur, or organic compounds and must be handled with the same precautions as explosives.

Analysis

Thermal decomposition of perchlorate salts to chloride, followed by the gravimetric determination of the resulting chloride, is a standard method of determining quantitatively the concentration of perchlorates. Any chlorates that are present in the original sample also break down to chloride. Thus results are adjusted to eliminate errors introduced by the presence of any chlorides and chlorates in the original sample.

The qualitative determination of water-soluble perchlorates by precipitation using methylene blue yields a violet precipitate (105). Using potassium, rubidium, or cesium salts for precipitation from ethanol–water solutions can serve as a qualitative determination of perchlorates (106). Tetraphenylarsonium chloride (107,108) has also been used for the precipitation of the perchlorate ion in gravimetric analysis.

Ion-specific electrodes can be used for the quantitative determination of perchlorates in the parts per million (ppm) range (109) (see ELECTROANALYTICAL TECHNIQUES). This method is linear over small ranges of concentration, and is best applied in analyzing solutions where interferences from other ionic species do not occur.

A practical method for low level perchlorate analysis employs ion chromatography. The unsuppressed method using a conductivity detector has a lower detectable limit of about 10 ppm. A suppression technique, which suppresses the conductivity of the electrolyte but not the separated ions, can further improve sensitivity (110,111). Additionally, ion chromatography can be coupled with indirect photometric detection and applied to the analysis of perchlorates (112).

Economic Aspects

Anhydrous perchloric acid is not sold commercially. Aqueous solutions of perchloric acid are sold at low concentrations for analytical standard applications and at concentrations up to 70%. The price for 70% perchloric acid varies and starts at \$2.70/kg, depending on the quantity and level of impurities.

The U.S. domestic capacity of ammonium perchlorate is roughly estimated at 31,250 t/yr. The actual production varies, based on the requirements for solid propellants. The 1994 production ran at about 11,200 t/yr, 36% of name plate capacity. Environmental effects of the decomposition products, which result from using solid rocket motors based on ammonium perchlorate-containing propellants, are expected to keep increasing public pressure until consumption is reduced and alternatives are developed. The 1995 price of ammonium perchlorate is in the range of \$1.05/kg. Approximately 450 t/yr of NH₄ClO₄-equivalent cell liquor is sold to produce magnesium and lithium perchlorate for use in the production of batteries (113). Total U.S. domestic sales and exports for sodium perchlorate are about 900 t/yr. In 1995, a solution containing 64% NaClO₄ was priced at ca \$1.00/kg; dry product was also available at \$1.21/kg.

Uses

Perchloric acid is used in analytical chemistry for the determination of trace metal constituents in oxidizable substances as well as in the production of high

purity metal perchlorates; it has also been introduced as a stable reaction media in the thermocatalytic production of chlorine dioxide (114). Perchlorates are primarily used in ammonium perchlorate as an oxidizer in the formulations of propellant for solid rocket motors. Perchlorates are used in the production of explosives, pyrotechnics, and in solid, slurried, and gelled blasting formulations. Both magnesium and lithium perchlorates are used in dry batteries. Other perchlorates have found application in oxygen-generation systems (qv) (115), adhesive bonding of steel plates (116), and the recovery of potassium from brines such as KClO₄ (see CHEMICALS FROM BRINE) (117,118).

BIBLIOGRAPHY

"Perchloric Acid and Perchlorates" under "Chlorine Compounds, Inorganic" in *ECT* 1st ed., Vol. 3, pp. 716–729, by H. L. Robson, Mathieson Chemical Corp., and J. C. Schumacher, Western Electrochemical Co.; under "Chlorine Oxygen Acids and Salts" in *ECT* 2nd ed., Vol. 5, pp. 61–84, by J. C. Schumacher and R. D. Stewart, American Potash & Chemical Corp.; in *ECT* 3rd ed., Vol. 5, pp. 646–667, by R. C. Rhees, Pacific Engineering & Production Co.

1. F. von Stadion, *Gilbert's Ann. Phys.* **52**, 197 (1816).
2. F. von Stadion, *Am. Chem. Phys.* **8**, 406 (1818).
3. R. C. Rhees, ed., *McGraw-Hill Encyclopedia of Science and Technology*, 2nd ed., McGraw-Hill Book Co., Inc., New York, 1966, p. 9.
4. J. J. Berzelius, *Lehrbuch der Chemie*, Section 2, Germany, 1835, p. 77.
5. C. W. Bennett and E. L. Mack, *Trans Am. Electrochem. Soc.* **29**, 323 (1916).
6. **U.S. Pat. 3,020,124** (1962), J. B. Brovo and P. H. Delano (to Foote Mineral Co.).
7. **U.S. Pat. 2,772,229** (1956), E. H. Karr (to Atochem North America).
8. **U.S. Pat. 5,131,989** (July 21, 1992), R. L. Dotson, R. L. Ralston, and H. J. Loftis (to Olin Corp.).
9. J. C. Schumacher, *Perchlorates, Their Manufacture & Uses*, ACS Monograph 146, Reinhold Publishing Corp., New York, 1960.
10. R. O. Segraves and D. Wickersham, *Chem. Eng. Prog.* **87**(6), 65 (1991).
11. **Swed. Pat. 8487** (1897), O. F. Carlson.
12. C. F. Goodeve and J. Powney, *J. Chem. Soc.*, 2078 (1932).
13. A. Pavia, *Compt. Rend.* **261**, 5118 (1965).
14. R. L. Fanar, *U.S. Atomic Energy Commission*, Report K1416, Washington, D.C., 1960.
15. C. F. Goodeve and F. D. Richardson, *J. Chem. Soc.*, 294 (1937).
16. R. Savoie and P. A. Giguere, *Can. J. Chem.* **40**, 990 (1962).
17. P. Jacobs, W. McCarthy, and H. M. Whitehead, *Chem. Rev.* **69**, 551 (1969).
18. K. Baum, *J. Org. Chem.* **41**, 1663 (1976).
19. A. A. Shilt, *Perchloric Acid and Perchlorates*, G. F. Smith Chemical Co., Columbus, Ohio, 1979.
20. F. C. Mathers, *J. Am. Chem. Soc.* **32**, 66 (1910).
21. H. H. Willard, *J. Am. Chem. Soc.* **34**, 1480 (1912).
22. D. Vorlander and R. von Schilling, *Liebig's Ann.* **310**, 369 (1900).
23. A. Michael and W. T. Conn, *Am. Chem. J.* **23**, 445 (1900).
24. T. W. Richards and H. H. Willard, *J. Am. Chem. Soc.* **32**, 4 (1910).
25. M. Volmer, *Liebig's Ann.* **440**, 200 (1924).
26. G. F. Smith, *Perchloric Acid*, 2nd ed., G. F. Smith Chemical Co., Columbus, Ohio, 1951.
27. K. C. Newman and F. C. Mathers, *Trans. Electrochem. Soc.* **75**, 271 (1939).
28. I. V. Shimonis, A. A. Rakov, and V. I. Veselovski, *Elektrokhimiya*, **6**, 169 (1970).
29. L. L. Bircumshaw and B. H. Newman, *Proc. R. Soc. London*, **A227**, 115 (1954).
30. *Ibid.*, 228 (1955).
31. *Ibid.*, 4741 (1957).
32. C. D. Hodgman, ed., *Handbook of Chemistry and Physics*, 37th ed., Chemical Rubber Publishing Co., Cleveland, Ohio, 1956.
33. S. Gordon and C. Campbell, *Anal. Chem.* **27**, 1102 (1955).
34. D. Vorlander and E. Kaascht, *Ber.* **56B**, 1157 (1923).
35. **U.S. Pat. 1,477,099** (Dec. 11, 1923), G. Baum.
36. R. L. Dotson, *J. App. Electrochem.* **23**, 897 (1993).
37. V. R. Pai Verneker and K. Rajishivar, *J. Solid State Chem.* **17**(1-2) 27 (1976).
38. J. W. Mellor, *A Comprehensive Treatise of Inorganic Chemistry*, Suppl. II, Pt. I, Longmans Green, New York, 1956.
39. K. A. Hofmann and A. Zedtwitz, *Ber. Dtsch. Chem. Ges.* **42**, 2031 (1909).
40. R. Binkley, Jr., *J. Am. Chem. Soc.* **62**, 3524 (1940).
41. N. V. Sidgwick, *Chemical Elements and Their Compounds*, Clarendon Press, Oxford, U.K., 1950.
42. F. Hein, *Chem. Tech. Berlin* **9**, 97 (1957).
43. G. F. Smith and V. R. Hardy, *Z. Anorg. U. Allgem. Chemie* **223**, 1–16 (1935).
44. H. H. Willard and G. F. Smith, *J. Am. Chem. Soc.* **45**, 286 (1923).
45. E. Hayck and E. Schnell, *Monatsb.* **85**, 472 (1954).
46. L. B. Serezhkina and co-workers, *Dokl. Akad. Nauk. SSR*, **211**(1), 123 (1973).
47. R. L. Berg and C. E. Vanderzee, *J. Chem. Thermodyn.* **73**, 219 (1975).
48. P. K. Jena and B. Prasad, *J. Indian Chem. Soc.* **31**, 480 (1954).
49. S. C. Sircar and B. Prasad, *J. Indian Chem. Soc.* **31**, 483 (1954).
50. E. Newbery, *Trans. Electrochem. Soc.* **69**, 611 (1936).
51. V. Krivtsov, V. P. Babaeva, and V. Ya Rosolovsky, *Zh. Neorg. Khim.* **18**, 353 (1973).
52. J. Gonzales de Barcia and E. Males, *Bol. Acad. Cienc. Exactas Fis. Nat. Madrid*, **2**, 8 (1936).
53. F. H. Spedding, J. L. Baker, and J. P. Watters, *J. Chem. Eng. Data*, **20**(2), 189 (1975).
54. V. Jedinakova, *Sb. Vys. Chem. Technol. Praze Anorg. Chem. Technol.* **B18**, 113 (1974).
55. H. Burton and P. F. G. Praill, *Analyst.* **80**, 4 (1955).
56. K. A. Hofmann, H. Kirmreuther, and A. Thal, *Ber. Dtsch. Chem. Ges.* **43**, 183 (1910).
57. K. A. Hofmann, A. Metzler, and H. Lecher, *Ber. Dtsch. Chem. Ges.* **43**, 178 (1910).
58. K. A. Hofman, R. Roth, K. Hobold, and A. Metzler, *Ber.* **43**, 2624 (1910).
59. V. P. Babaev and V. Y. Rosolovskii, *Izv. Akad. Nauk SSSR, Ser. Khim.* (11) 2414 (1974).
60. R. Salvadori, *Gazz. Chim. Ital.* **37**(II), 32 (1907).
61. D. R. Goddard, E. D. Hughs, and C. K. Ingold, *J. Chem. Soc.* 2559 (1950).
62. W. R. Angus and A. H. Leckie, *Proc. R. Soc. London*, **A-150**, 615 (1935).
63. L. J. Klinkenberg, *Rec. Trav. Chim.* **56**, 749 (1937).
64. K. Cruse and co-workers, *Z. Anorg. Chem.* **259**, 154 (1949).
65. E. J. Arlman, *Rec. Trav. Chim.* **56**, 919 (1937).

66. U.S. Pat. **4,012,492** (1977), C. J. Schack and D. Pilipovich (to U.S. Dept. of the Navy).
67. G. F. Potapoba, A. A. Rakov, and V. I. Vesselovskii, *Elektrokhimiya*, **9**, 1054 (1973).
68. W. Oechsli, *Z. Electrochem.* **9**, 807 (1903).
69. R. D. Stewart, in J. C. Schumacher, ed., *Perchlorates*, Reinhold Publishing Corp., New York, 1960, Chapt. 5.
70. S. V. Loginov and co-workers, *Khim. Kompleksn. Soedin.* **3**(12), 389 (1975).
71. R. E. Elson and J. E. Stucky, *J. Inorg. Nucl. Chem.* **35**, 1029 (1973).
72. V. E. Kalinina and co-workers, *Khim. Tekhnol.* **19**(8), 1287 (1976).
73. A. E. Wickenden and R. A. Krause, *Inorg. Chem.* **4**, 404 (1965).
74. J. Lewis, R. S. Nyholm, and G. A. Rodley, *Nature*, **207**, 73 (1965).
75. W. R. McWhinnie, *J. Inorg. Nucl. Chem.* **26**, 21 (1964).
76. H. C. Clark and R. J. O'Brien, *Inorg. Chem.* **2**, 740 (1963).
77. G. S. Serullas, *Ann. Chem. Phys.* **45**, 270 (1830).
78. N. A. E. Millon, *Ann. Chim. Phys.* **7**, 298 (1843).
79. H. Kolbe, *Ann. Chem.* **64**, 237 (1847).
80. T. Fairley, *Br. Assoc. Adv. Sci. Rep.* **44**, 57 (1874).
81. S. Vaclav, S. Tichy, and A. Regner, *Chem. Prum.* **16**, 577 (1966).
82. U.S. Pat. **2,392,861** (Jan. 15, 1946), J. C. Pernert (to Oldbury Electrochemical Co.).
83. W. Möller and P. Jünck, *Chem. Ing. Tech.* **35**, 78 (1963).
84. U.S. Patent **5,160,416** (Nov. 3, 1992), D. W. Cawfield and co-workers (to Olin Corp.).
85. U.S. Patent **5,294,319** (Mar. 15, 1994), J. J. Kaczur and co-workers (to Olin Corp.).
86. U.S. Patent **5,298,280** (Mar. 29, 1994), J. J. Kaczur and co-workers (to Olin Corp.).
87. W. Oechsli, *Z. Electrochem.* **9**, 807 (1903).
88. N. V. S. Knibbs and H. Palfreeman, *Trans. Faraday Soc.* **16**, 402 (1920).
89. K. Sugino and Aoyagi, *J. Electrochem. Soc.* **103**, 166 (1956).
90. C. W. Bennett and E. L. Mack, *Trans. Electrochem. Soc.* **29**, 323 (1916).
91. M. P. Grother and E. H. Cook, *Electrochem. Technol.* **6**, 221 (1968).
92. E. H. Cook and M. P. Grother, paper presented at *The Electrochemical Society Meeting*, Los Angeles, Calif., May 1971.
93. U.S. Pat. **3,475,301** (Oct. 28, 1969), E. H. Cook and M. P. Grother (to Hooker Chemical Corp.).
94. A. Legendre, *Chem. Ing. Tech.* **34**, 379 (1962).
95. E. Hausmann and E. Kramer, *Chem. Ing. Tech.* **43**, 170 (1971).
96. *Chem. Eng.* **72**, 82 (July 19, 1965).
97. U.S. Pat. **2,945,691** (July 19, 1960), F. D. Gibson, Jr. (to Pacific Engineering & Production Co.).
98. U.S. Pat. **3,634,216** (Jan. 11, 1972), F. D. Gibson, Jr., R. L. Thayer, and B. B. Halker (to Pacific Engineering & Production Co.).
99. K. C. Narasimhaw and H. V. K. Udupa, *J. Electrochem. Soc.* **123**, 1294 (1976).
100. M. Nagalingam and co-workers, *Chem. Ing. Tech.* **41**, 1301 (1969).
101. U.S. Pat. **3,493,478** (Feb. 3, 1970), H. V. K. Udupa and co-workers (to Central Electrochemical Research Institute, India).
102. J. C. Schumacher and D. R. Stern, *Chem. Eng. Prog.* **53**, 428 (1957).
103. J. E. Reynolds and T. W. Clapper, *Chem. Eng. Prog.* **57**(11), 138 (1961).
104. C. L. Keller, *Hazardous Materials Regulations, Tariff No.* BOE, U.S. Department of Transportation, Washington, D.C., Apr. 23, 1992.
105. F. D. Snell and C. T. Snell, *Colorimetric Methods of Analysis*, Vol. II, D. Van Nostrand Co., Princeton, N. J., 1957, pp. 718–719.
106. H. H. Willard and H. Diehl, *Advanced Quantitative Analysis*, D. Van Nostrand Co., Princeton, N.J., 1944, pp. 254–257.
107. G. M. Smith, *Ind. Eng. Chem. Anal. Ed.* **11**, 186 (1939).
108. *Ibid.*, 269 (1939).
109. R. J. Bacquik and R. J. Dubois, *Anal. Chem.* **40**, 685 (1968).
110. H. Small and T. E. Miller, Jr., *Anal. Chem.* **54**, 462 (1982).
111. U.S. Pat. **4,474,664** (Oct. 2, 1984), T. S. Stevens, J. C. Davis, and H. Small (to Dow Chemical Co.).
112. U.S. Pat. **4,414,842** (Nov. 15, 1983), H. Small and T. E. Miller, Jr. (to Dow Chemical Co.).
113. Technical data, Kerr-McGee Chemical Corp., Oklahoma City, Okla., May 1994.
114. K. E. Woodard, Jr., D. W. Cawfield, and S. K. Mendiratta, "Chloric Acid: A New Electrochemical Product," *183rd Electrochemical Society Meeting*, Honolulu, Hawaii, May 1993.
115. U.S. Pat. **3,993,514** (Nov. 23, 1976), E. J. Pacanowsky and E. A. Martino (to Thiokol Corp.).
116. U.S. Pat. **3,993,524** (Nov. 23, 1976), Y. Okada and S. Kensho (to Nissan Motor Co., Ltd.).
117. J. A. Epstein and co-workers, *Hydrometallurgy*, **1**, 39 (1975).
118. D. R. George, J. M. Riley, and J. R. Ross, paper presented at *The 62nd National Meeting of the Institute of Chemical Engineers*, Salt Lake City, Utah, May 21–24, 1967.

Sudhir K. Mendiratta
Ronald L. Dotson
Robert T. Brooker
Olin Corporation

PERCHLORO COMPOUNDS.

See CHLOROCARBONS AND CHLOROHYDROCARBONS.

PERFLUORINATED IONOMER MEMBRANES.

See IONOMERS; MEMBRANE TECHNOLOGY.

PERFLUORO COMPOUNDS.

See FLUORINE COMPOUNDS, ORGANIC.

PERFUMES

Perfumes are mixtures created for use in a wide variety of applications, ranging from expensive couturier perfumes to cosmetics, personal grooming products, laundry products, household cleaning products, and many others. They are created from a palette of several thousand materials, most of which are manufactured by chemical processing methods. Until late in the nineteenth century, fragrances were derived from natural sources, which put a limitation on where and how they could be used. The large increase in the use of perfumes since then would not have been possible without the chemical developments that allowed the synthesis and commercial production of many new odorants. As a result, fragrances, once a luxury, have been incorporated routinely into a great number of products in daily use.

Historical Background

The use of fragrance, deeply rooted in human experience, predates written history. The earliest evidence of such activity is archaeological, found in the tombs of the First Dynasty Egyptian Kings (5100–3500 BC). Small alabaster vases found along with personal grooming tools are taken as an indication of the use of perfumes and cosmetics. Such vases found in the tomb of King Tutankhamen (1350 BC) in 1926 contained some oily material which had remained elusively fragrant. Analytical methods used at that time were able to indicate only that the material appeared to be made mostly of animal fat and some sort of balsamic substance. Early indications of perfume-making, dating back to 3000 BC, have been found in the former Mesopotamia. Extraction pots for herbal and fragrance preparations found there may have been the world's oldest distillation apparatuses.

The ancients used fragrances for a variety of reasons. Perfumes were offerings to the gods, probably in the form of incense; they were used for aesthetic purposes as part of personal grooming in daily life; and they were used in Egyptian embalming rituals. It has been suggested that priests were the first perfumers. There are many Old Testament references to perfume ("anointing with oils"), beginning mainly in the book of Exodus. Thus it seems likely that the ancient Hebrews learned to use fragrance from the Egyptians.

Among the earliest writings on compounded perfumes, ie, deliberately created mixtures rather than simple extracts, are those of the Greek, Theophrastus, from 370 BC. He even referred to the use of oil bases to make perfumes long-lasting, something that remains a challenge to the modern perfumer. Although such writings typically focus on Western history, there are ancient writings and artifactual evidence from India and China indicating that the use of perfumes and cosmetics developed independently in those countries, where fragrance use is traditional in the cultures.

By the thirteenth century AD, essential oils were being produced along with medicinal and herbal preparations in pharmacies. Around this time improvements in distillation techniques were made, in particular the development of the alembic apparatus, which would eventually establish the quality of such materials. As a result, many of the essential oils in use today are derived from those produced in the sixteenth and seventeenth centuries in terms of odor character, even though production methods have continued to evolve. The current practice of aroma therapy is an indication of this common root of medicinal and fragrance chemistry.

During the middle of the nineteenth century, chemists began to investigate the compositions of natural fragrance and flavor materials. As the science developed, it became possible to identify and then synthesize many specific chemicals; this was followed by the use of chemicals as perfume ingredients, first as materials isolated from essential oils and then as synthetically produced naturals such as vanillin and coumarin [91-64-5]. Late in the nineteenth century, it was discovered that synthetic materials that are unrelated to natural chemicals could have great value as perfume ingredients; nitro musks and the ionones were among the first such materials found. Since that time, synthetic chemicals have grown steadily in numbers and in total use compared with materials of natural origin.

Creation of Perfumes

Perfumes are usually considered in two broad categories, as either fine or functional (household product) fragrances. Fine fragrances include perfumes, colognes, men's colognes, aftershaves, and fragrances for cosmetic products. For the purposes of this article, functional products include all personal and household cleaning products that are perfumed, such as bar soaps, detergents of various types, fabric softeners, and bleach formulations. The technical and economic requirements for various kinds of perfumes differ widely and are taken into account during their creation. Creation of a successful fragrance can be a lengthy, painstaking, and costly process. The investment in each fragrance that reaches the marketplace is quite substantial. For these reasons, fragrance formulas are held as trade secrets. Patent protection is generally not suitable for this type of intellectual property because most perfume components are in the public domain; it is the unique combination of ingredients that affords competitive advantage. The formulas are significant assets of the companies that produce them and are therefore protected by elaborate security arrangements.

The creation of perfumes is a commercial art practiced in the medium of odorous substances; it is highly specialized and individualistic. A wide variety of ingredients, numbering in the thousands and varying greatly in chemical composition, is available to perfumers. The choice of ingredients can be based on aesthetic or technical grounds, depending on the intended use of the perfume. Often a fragrance creation begins with a concept inspired by an existing perfume, a newly available ingredient, or a newly discovered odor facet of an existing material. The perfumer creates an accord based on the inspiring note. The new accord may be used in an existing floral or woody-floral fragrance base, or a new floral composition may be created around the original theme. Computer-assisted design methods are utilized by perfumers to help with formulation changes and to keep track of the large material base

available to them.

The performance of a fragrance over time in its intended application is usually an important consideration. A perfume can be viewed as a blend with a top note continuing into a middle and on to an end note. As might be expected, this is a function of relative volatility and odor strengths of the materials used. The perfumer must smooth the odor profile of a formula so that there are no discontinuities in odor impact as the different components evaporate. The amount of an ingredient in a fragrance formulation can vary widely, from 10–20^o for some materials to trace levels (parts per million) for others. The cost of an ingredient or its odor strength can be a limitation on its use.

Perfumes and ingredients are evaluated initially by a simple bioassay: smelling them from paper strips or blotters. However, this apparently straightforward exercise is made extremely complex by the ability of human beings to perceive hundreds of thousands of different odor nuances, by individual differences in odor perception, and by differences in the way individuals describe odors. The phenomenon of fatiguing or adaptation to odors is also a complicating factor, which is often not taken into account. The simple criterion of a material smelling good or pleasant is entirely inadequate to determine its value as an odorant. In fact, some materials, both of natural and synthetic origin, smell rather unpleasant and yet are important in perfumery. Even after years of experience in the hands of many perfumers, new ways to use a particular fragrance ingredient can be discovered and become the basis of a new fragrance type.

Odor perception and description are highly subjective in nature. Nevertheless, there is a generally agreed-upon odor vocabulary that is used to characterize individual ingredients and finished fragrances. Table 1 shows some commonly used odor descriptors grouped into five general classifications.

Table 1. Perfumery Descriptions

	Floral	Citrus	Woody	Green	Fruity
carnation	lilac	bergamot	cedar	basil	apple
chrysanthemum	lily	grapefruit	fir	cucumber	apricot
gardenia	marigold	lemon	hickory	grass	banana
honeysuckle	muguet	lime	patchouli	parsley	black currant
hyacinth	narcissus	mandarin	pine	rhubarb	cherry
iris	orange flower	orange	sandal	string bean	fig
jasmine	rose	tangerine		violet	grape
jonquil	violet	verbena		watercress	melon
lavender	mimosa				peach
					pineapple
					prune
					raspberry
					strawberry

Fine Fragrances

Fine fragrances must work on the skin and blend with body odor. They must be pleasant, diffusive, and substantive (long-lasting), and have the quality of genuine beauty and signatures that distinguish them from each other. For most fine fragrances, the perfumes are themselves the products. They are sold to the consumer at various concentrations in alcoholic or aqueous–alcoholic solutions, depending on the type of application intended. For example, women's perfumes are typically 20–35^o fragrance oil in 95^o ethanol. Women's colognes are offered in the range of 15–22^o fragrance oil, whereas men's colognes and aftershaves are usually in the range of 2–12^o. The creation of fine fragrances allows for the highest degree of freedom in terms of ingredient choice and economics. Consequently, fine fragrances often set trends that eventually find their way into other kinds of products. Perfumes can be grouped into broad odor categories in an attempt to show their relationships to each other and sometimes indicate the progress of creative evolution as new fragrances are built on the foundations laid by older ones. Following are a number of fine fragrances grouped by a widely used classification scheme.

Women's Fragrances.

Straight Floral Family. The straight floral family contains a large and popular group of flowery odors, most of them easily recognizable.

Carnation

Bellodgia (Caron 1927)
Spellbound (Lauder 1992)

Jasmine

Honeysuckle (Avon 1963)
Dior Dior (Dior 1976)
Chevrefeuille (Rocher 1976)

Rose

Tea Rose (Workshop 1972)
Evelyn Rose (Crabtree & Evelyn 1993)

Muguet

Diorissimo (Dior 1956)
Sunflower (Arden 1993)
Diamonds & Sapphires (Arden 1993)
Tuberose
Fracas (Piguet 1945)

Floral Bouquet Family.

In the floral bouquet family, fantasy accords are blended into the floral. They all have distinct notes that distinguish one perfume from another.

White Shoulders (tuberose)

White Shoulders (Evyan 1939)
Chloe (Lagerfeld 1975)
Amarige (Givenchy 1992)

Fidji (floral, green)

Fidji (Laroche 1966)
Norell (Revlon 1969)
Charlie (Revlon 1973)

Joy (rose, jasmin, muguet)

Joy (Patou 1930)
First (Van Cleef & Arpels 1977)
White Linen (Lauder 1978)
L'Air du Temps (spicy carnation)
L'Air du Temps (Ricci 1948)
Paris (Yves Saint Laurent 1984)
Eternity (Calvin Klein 1988)

Aldehydic Floral Family.

This is an important family of fragrances, the typical odor of which is the class odor of the aldehydes. The aldehydes are present in small quantities in nature and have an unnatural brilliance. Although they have sharp, slightly fruity odors alone, they blend beautifully and unexpectedly in florals.

Chanel No. 5 (floral, aldehydic)

Madame Rochas (woody, mossy, peach)

Chanel No. 5 (Chanel 1921)
L'Interdit (Givenchy 1957)
Delicious (Gayle Hayman 1993)

Madame Rochas (Rochas 1960)
Calandre (Paco Rabanne 1970)
Infini (Caron 1970)
Nude (Bill Blass 1990)

Oriental Family. In these perfumes, a mossy, woody, and spicy accord combines with the sweetness of vanilla or balsam and is accented with animal notes such as amber, civet, and musk. The most important floral accords used are rose and jasmine.

Oriental
Youth Dew (Lauder 1953)
Opium (YSL 1977)
Cinnabar (Lauder 1978)
Sweet vanilla
Emeraude (Coty 1921)
Shalimar (Guerlain 1925)
Exclamation (Coty 1988)
Tresor (Lancome 1991)

Orange flower spice
Aprus L'Ondine (Guerlain 1906)
L'Origan (Coty 1909)
Private Collection (Lauder 1973)
Oscar de la Renta (de la Renta 1977)
J'ai Osé (Laroche 1978)

Chloé Narcisse (Parfums International Ltd. 1992)
Volupté (Oscar de la Renta 1992)

Chypre Family. The fragrances of this large and important group are warm, mossy, and long-lasting, having rose, jasmine, and animal notes. By blending different accords in the chypre (moss) base, a large new fragrance group is created. Accords that blend well are fruity, green galbanum, aldehydic, and leathery in character.

Chypre
Chypre (Coty 1917)
Parure (Guerlain 1975)
Oriental
Pavilion (Lauder 1978)
Mystère de Rochas (Rochas 1978)
Bandit (woody amber)
Bandit (Piguet 1944)
Cabocharde (Grès 1958)
Diva (Ungaro 1961)
Ungaro (Ungaro 1977)
Miss Dior (chypre patchouli aldehyde green)
Miss Dior (Dior 1947)
Givenchy III (Givenchy 1970)
Ysatis (Givenchy 1985)

Halston (amber, woody)
Halston (Halston 1975)
Red (Giorgio 1989)
Mitsouko (chypre peach)
Mitsouko (Guerlain 1919)
Femme (Rochas 1945)
Diorélla (Dior 1972)
Champagne (YSL 1993)
Crope de Chine (chypre aldehyde)
Crope de Chine (Milot 1928)
White Diamonds (Arden 1992)

Woody Family. The perfumer has available many different woody fragrance materials, both natural and synthetic. Naturals such as sandal, vetivert, cedar, and patchouli often form the bases of these fragrances. They combine in harmony with sweet notes, florals, and animal accords.

Orris
Chamade (Guerlain 1970)
Chanel No. 19 (Chanel 1971)
Calvin Klein (Klein 1978)
Safari (Ralph Lauren 1990)

Patchouli
Shocking (Schiaperelli 1935)
Clinique Aromatics (Lauder 1971)
Knowing (Lauder 1988)

Green Family. This group has taken a long time to gain widespread acceptance, although among perfumers, green notes have always enjoyed great popularity.

Vent Vert (Balmain 1945)
Aliage (Lauder 1972)
Cristalle (Chanel 1974)

Lauren (Lauren 1978)
Escape (Calvin Klein 1991)
L'eau D'Issey (1992)

Citrus Family. This blend has always been popular and in the 1980s experienced a revival in which citrus was blended with florals and sweet notes.

Jean Marie Farina (Roger & Gallet 1806)
Φ De Lancôme (Lancôme 1975)

Quartz (Molyneux 1977)
Calyx (Lauder 1986)

Musk Family. Even though the musk trend began as a counterculture fashion, these accords have become widely accepted.

Musk Oil (Caswell-Massey 1950)
Musk Oil (Jovan 1972)

Wild Musk (Coty 1993)

Leather Family. Fragrances of this group can always be recognized by the warm, leathery tobacco note.

Tabac Blond (Caron 1919)
Cuir de Russe (Chanel 1924)

Donna Karan (Donna Karan Beauty Co. 1993)

Men's Fragrances. Earlier in the twentieth century, men's fragrances were expected to have a masculine direction, such as tobacco, leather, fougere, or citrus, even if only in name. This is no longer true, however; since the 1970s, men's perfumes have become less conservative and have allowed much more creative use of rich woody, ambery, and green notes.

Green Family. A relatively easily recognizable group having definite green character, the green family is increasingly popular, but is still regarded as rather exclusive.

Green
Old Spice Herbal (Shulton 1974)

Herbal
Grey Flannel (Beene 1975)
Devin (Lauder 1977)
Fahrenheit (Dior 1989)

Citrus Family. This is a popular fragrance group noted for its refreshing brisk quality. Lemon, lime, orange, and bergamot are important ingredients. These oils combine well with lavender and amber accords.

Lavender
English Lavender (Yardley 1770)
Pour un Homme (Caron 1934)
Eternity (Klein 1988)
Coolwater (Davidoff 1991)

Pine
Pino Silvestre (Vidal 1948)
Aqua di Silva (Victor 1948)
Eau Sauvage
Eau Sauvage (Dior 1966)
Bravas (Shiseido 1969)

Fougere Family. This family has a typically accepted masculine note reminiscent of fern, tonka, and moss.

Fougere Royal (Houbigant 1822)
Jicky (Guerlain 1889)

Moustache (Rochas 1949)
Drakkar Noir (Guy Laroche 1984)

Canoe Family. This is one of the most popular fragrance families dating back to the 1940s. Also liked by women, it has a typical unisex note.

Canoe (Dana 1935)

Brut (Faberge 1964)

Spice Family. This is an easily recognizable fragrance that has a strong, spicy character, eg, Old Spice (Shulton 1937).

Woody Family. This group of fragrances is well appreciated by men. It owes its character to woody naturals, such as vetivert and patchouli, but has become more complex over the years owing to the use of a large variety of new woody aroma chemicals that gave their own signature to the later fragrances in the group.

Amber
Halston Z-14 (Halston 1976)
Yatagan (Caron 1976)
Adventurer (Bauer 1993)
Sandalwood
Arden for Men (Arden 1955)
YSL (YSL 1971)

Patchouli
Aramis 900 (Lauder 1970)
Monsieur Jovan (Jovan 1975)
Polo (Ralph Lauren 1978)
Egoiste (Chanel 1991)

Musk Family. Musk notes, combined with previously accepted accords, have made this a popular group.

Musk for Men (Yardley 1971)
Musk for Men (Jovan 1973)

Old Spice Musk for Men (Shulton 1974)
Royal Copenhagen Musk (Swank 1975)

Leather Family. A more or less leathery character is typical of this group.

Knize Ten (Caswell-Massey 1927)
Ted (Lapidus 1978)

Bel Ami (Hermes 1987)

Oriental Family. Sweet, balsamic notes are typical for this group.

Habit Rouge (Guerlain 1964)
Pierre Cardin (Cardin 1972)

Lagerfeld (Lagerfeld 1978)

Chypre Family. This is an extremely popular group that was well received at the end of the 1940s. It is not exclusively masculine since many women's perfumes have been used as models for fragrances in this family. This group represents creativity at its best and shows how well a basic chypre accord can blend with new themes.

Zizanie (Fragonard 1932)
Royal Copenhagen (Swank 1971)
Paco (Paco Rabanne 1973)
Revlon Pour Homme (Revlon 1977)

Givenchy Gentlemen (Givenchy 1971)
Aramis (Lauder 1965)

Azzaro (Azzaro 1978)

Functional Fragrances

Perfumes for functional products are used in different ways than are fine fragrances. The most obvious difference is that functional fragrances are incorporated into a variety of media in relatively small amounts. Perfuming of cleaning products probably began with efforts to cover the undesirable odors that accompanied the rendering of tallow to make soap. Although consumers have come to enjoy pleasant-smelling personal and household cleaning products, the covering of malodors in the product bases is still a significant challenge to the perfumer (see *ODOR MODIFICATION*). Product bases may also contain ingredients that react with certain perfume ingredients to alter or destroy their odors, or to cause discoloration problems. Thus the task of the functional products perfumer is dominated from the outset by the nature and economics of the product to be fragranced. Aroma chemicals have important advantages over essential oils and other naturals in functional applications because the former are much better characterized in terms of chemical type and reactivity. A knowledge of the chemical properties of fragrance materials is of great value to the functional products perfumer. However, it has been found rather difficult to make precise predictions of material stabilities because product bases are frequently changing and their complete compositions are generally not revealed. Therefore, a certain amount of empirical testing in product bases is necessary. The following are brief descriptions of the fragrance requirements for a number of household products and the perfumery approaches used for them.

Detergent Fragrances. There is a great variety of laundry detergent formulations on the market, and new ones are introduced every year. The incorporation of bleaching agents into laundry products and the advent of highly concentrated detergents present new and increasingly difficult challenges to the perfumer. Several factors play important roles. Most critical are chemical stability of the fragrance material in detergent and the rate of evaporation from the sales package. Also important is the performance in the product's end use, ie, in the wash water and on the laundered cloth. To some degree, these factors can be predicted from chemical principles and physical properties, but testing of individual materials under use conditions is also important. Chemical stability can be checked by incorporating the fragrance material in the detergent base packed in a closed container and subjecting it to accelerated aging, eg, 40°C for one month. Storage tests in cardboard containers under controlled, standardized temperature and humidity give a good picture of both chemical and evaporative performance. The country of destination is important in determining stability test conditions. Clearly the customer requirements for laundry detergents in Canada are different from those in Brazil.

Detergent fragrances must be particularly powerful and effective because they are incorporated into the final product at rather low levels. Typical detergent powders can contain as little as 0.3% fragrance, although this may be higher in concentrated products. The fragrance should cover undesirable odors in the product itself and those resulting from soils in the wash solution. It should also reinforce the performance of the laundry product by imparting a pleasant scent to the clean and dried fabrics. Substantive ingredients such as Galaxolide, Lyrall, Lilial, and Ambroxan are used to obtain residual fragrance on the cloth.

Soap Fragrances. The function of soap is to clean; however, the fragrance, at a dosage of 1–2%, plays a large role in the perceived quality of the soap bar. Aesthetically, a beauty soap requires a different type of fragrance from a deodorant soap or a freshness bar. The odor types can be single florals or complicated perfumes in which fine fragrances have been the models for the soap creation. Besides the aesthetic quality of the soap fragrance, there are a number of technical complications that the perfumer has to deal with: limitations on cost, odor quality and other characteristics of the soap base, the presence of additives, and the high pH of most soaps (often between 9.5 and 11.0) which may lead to hydrolysis or discoloration problems. The main ingredients in bar soaps are derived from a variety of sources, ie, animal and vegetable fats as well as synthetics, which result in differing demands on fragrance performance. In addition to predicting the performance and stability of individual ingredients, in many cases the perfumer must test a large number of materials, each separately, in the soap base. As with detergents, the test conditions can vary according to the requirements of the customer and the country where the soap is to be marketed. For example, the soap bar in a typical package can be stored in an oven at 40°C for one month or in a glass jar at 60°C for seven days. The perfumer's interest is in the performance of each material in the soap base at room temperature, stability after accelerated tests, and coloration or possible discoloration in light and at elevated temperatures. Armed with this information for a large number of ingredients and a creative idea, the perfumer can work toward the optimal soap fragrance. The final fragrance must be subjected to stability tests.

Liquid Fabric Softeners. The principal functions of fabric softeners are to minimize the problem of static electricity and to keep fabrics soft (see *ANTISTATIC AGENTS*). In these laundry additives, the fragrance must reinforce the sense of softness that is the desired result of their use. Most fabric softeners have a pH of about 3.5, which limits the materials that can be used in the fragrances. For example, acetals cannot be used because they break down and cause malodor problems; in addition, there is the likelihood of discoloration from Schiff bases, oakmoss extracts, and some specialty chemicals. Testing of fragrance materials in product bases should take place under accelerated aging conditions (eg, 40°C in plastic bottles) to check for odor stability and discoloration.

A special requirement of perfumes for fabric softeners is the ability to leave a residual of odor on fabric after line- or machine-drying. Substantivity of different fragrance materials can be determined through empirical testing or by analytical methods. Unfortunately, increasing the percentage of substantive materials in the fragrance is not always a guarantee of a good fabric softener fragrance. This is because substantive materials must not only blend into the fragrance type as well as any odor from the product base, but they must also survive rinsing. The information gained through testing, however, can help the perfumer choose the right fragrance type and include a fair amount of substantive materials.

Tumble-Dryer Softeners. In these products, which are designed for machine drying, the carrier, in the form of a nonwoven fabric, a foamed plastic, or a cotton string, contains the active softener ingredient and the fragrance. The fragrance partly disappears with the hot drying air and is partly absorbed into the fabrics. Testing for this application requires many drying cycles to obtain substantivity information on a range of individual fragrance materials. Aesthetically, the fragrance should support a sense of softness and caring for fine laundry.

Bleach Products. Many marketed bleach products, especially those containing hypochlorite, are still unperfumed. This is understandable because hypochlorite bleaches impose severe limitations on the fragrance materials that can be used due to their oxidizing power and high pH of 12.5. The same factors also limit the solubilizers that can be used. These problems have been largely overcome through knowledge of the chemical properties of available aroma chemicals and the selective development of new ones. Fragrances for bleaches are readily created to support the powerful cleaning properties of such products.

Shampoo Perfumes. The stability of perfume in shampoo is usually not a problem because shampoo pH is near neutral; it is only when special additives are used that stability and performance tests may be required. However, in some cases the addition of perfume can affect properties of the shampoo such as viscosity. Testing is therefore often necessary to identify interactions that can negatively affect the product. The fragrance is often required to support the concept and image of the product, eg, in herbal or balsam shampoos. Fragrance dosages are generally 0.5–1.0% for normal shampoos, but can be up to 1.5% when a particularly strong odor and residual effect are desired.

Deodorants and Antiperspirants. Deodorants are often made in sticks or lotions. Sticks are mainly alcoholic or glycolic soap gels. Fragrances for these products have stability requirements similar to soap fragrances, and must also withstand the relatively high temperatures (~60°C) encountered during manufacture. Depending on the formulation of the lotion, existing fragrances may have to be modified. Traditional deodorants use bacteriostats to help lower body odor. Antiperspirants usually contain aluminum or zirconium salts that can reduce the pH to about 2.3, which makes acid-sensitive fragrance materials unsuitable for this application.

Deodorant fragrances must be long-lasting in order to help maintain a pleasant body odor for as long as possible. In the 1990s when naturalness is the trend, a certain amount of fresh sweat odor is acceptable. In a modern approach to deodorants, the body odor is used as an animal accord that blends well with the rest of the fragrance. Thus the perfumer attempts to create a fragrance that smells good both in its original form and with the addition of the body odor as it develops.

Talcs and Powders. In the perfuming of talcum powders and face powders, stability is the most important factor (see also *COSMETICS*). Even the finest talc contains alkaline impurities that can cause decomposition and discoloration of fragrance ingredients. In addition, the fragrance is spread thinly around the microscopic talc particles, accelerating oxidative reactions. For this reason, many aldehydes and terpenes fare poorly and vanish completely as odor contributors. Systematic screening programs are often necessary to obtain a list of stable fragrance raw materials for these applications.

Perfume Ingredients

The classical materials of perfumery are natural products. These are mostly of vegetative origin, with some obtained from animal secretions. Just about any part of a plant can be used, including flowers, fruits, leaves, twigs, roots, and wood, depending on the amount and quality of essence it contains. Perfume materials of animal origin include tincture of tonquin musk, civet gum, and beaver castoreum. Some such materials have been or are being replaced by synthetic substitutes for environmental, political, or economic reasons. Even though synthetics continue to grow and have displaced naturals in overall usage, the latter are deeply rooted in the art of perfumery and so remain extremely valuable in perfume creation; in addition, naturals provide the odor reference points by which synthetics are often judged or described.

Natural Products. Various methods have been and continue to be employed to obtain useful materials from various parts of plants. Essences from plants are obtained by distillation (often with steam), direct expression (pressing), collection of exudates, enfleurage (extraction with fats or oils), and solvent extraction. Solvents used include typical chemical solvents such as alcohols and hydrocarbons. Liquid (supercritical) carbon dioxide has come into commercial use in the 1990s as an extractant to produce perfume materials. The principal forms of natural perfume ingredients are defined as follows; the methods used to prepare them are described in somewhat general terms because they vary for each product and supplier. This is a part of the industry that is governed as much by art as by science.

Concretes. Concretes are produced by extraction of flowers, leaves, or roots, usually with hydrocarbon solvents. After removal of the solvent by distillation, the concrete is obtained as a thick, waxy residue. Such materials are used in some fine fragrances, but the waxes they contain can give rise to solubility problems. For this reason, concretes are often dissolved in alcohol to make tinctures, or in other low odor diluents. Production of concretes, especially flower concretes, usually takes place where the botanicals are grown since the odors of such materials deteriorate rapidly after harvesting.

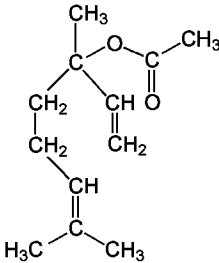
Absolutes. Absolutes are prepared from concretes by further processing to remove materials that can cause solubility problems in perfumes. This is done by dissolution in alcohol, filtering, and removal of the solvent, usually at reduced pressures. The resulting products are viscous, oily materials which may be diluted with low odor substances such as diethyl phthalate.

Concretes and absolutes, both obtained by total extraction of the plant material and not subject to any form of distillation other than solvent removal, are complex mixtures containing many chemical types over wide molecular weight ranges. In some cases, gas chromatographic analysis shows little volatile material. Yet these products have powerful odors and contribute in important ways to the perfumes in which they are used.

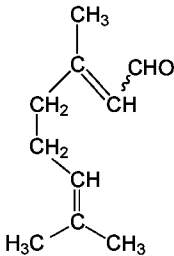
Essential Oils. Essential oils are produced by distillation of flowers, leaves, stems, wood, herbs, roots, etc. Distillations can be done directly or with steam. The technique used depends mostly on the desired constituents of the starting material. Particular care must be taken in such operations so that undesired odors are not introduced as a result of pyrolytic reactions. This is a unique aspect of distillation processing in the flavor and fragrance industry. In some cases, essential oils are obtained by direct expression of certain fruits, particular of the citrus family. These materials may be used as such or as distillation fractions from them (see OILS, ESSENTIAL).

Naturally Derived Materials. The following are descriptions of some of the most important naturally derived materials in use. Importance in this context is defined in terms of the total value of the materials, which range from expensive, low volume materials that have great aesthetic value to relatively inexpensive and widely used products. For some of the naturals, it is indicated whether they can be distilled to provide individual chemicals for use as such or as intermediates. Materials produced in this way from a given natural source are usually not interchangeable with those from other naturals or synthetics. In some cases this may be due to optical isomerism, which can have a significant effect on odor, but usually it is due to trace impurities.

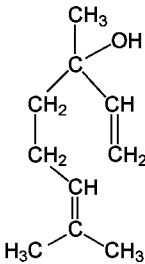
Bergamot. Bergamot oil is produced by cold expression from peels of fruits from the small citrus tree, *Citrus bergamia*. The fruits are inedible and of little value. Bergamot is grown mainly in southern Italy and northern and western Africa. The oil is used to impart a sweet freshness to perfumes. Its largest chemical constituent, to the extent of 35–40%, is linalyl acetate [115-95-7] (1), with a much smaller amount of citral [5392-40-5] (2) as an important odor contributor.



(1)



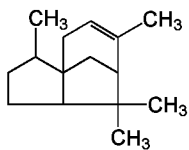
(2)



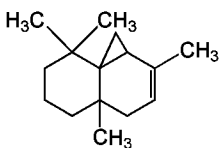
(3)

Bois de Rose. Bois de rose oil is obtained by steam distillation of wood chips from South American rosewood trees, *Aniba msaenodora*. The tree, a wild evergreen, grows mainly in the Amazon basin. The oil is used as obtained in perfumery for its sweet, woody-floral odor and as a source of linalool [78-70-6] (3), which it contains to the extent of 70%. Linalool distilled from bois de rose oil is also used directly in perfumery and for conversion to esters, eg, the acetate (1).

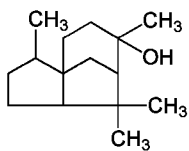
Cedarwood. Many varieties of cedarwood oil are obtained from different parts of the world. They are produced mainly by steam distillation of chipped heartwood, but some are also produced by solvent extraction. The oils, which vary significantly in chemical composition, are used in perfumes as such, but the main uses are as distillation fractions and chemical derivatives. For the latter purposes the most used oils, which are similar in composition, are from Texas in the United States (*Juniperus mexicana*) and from China (*Cupressus funebris*). The principal constituents of these oils are cedrene [11028-42-5] (4), thujopsene [470-40-6] (5), and cedrol [77-53-2] (6). The first two of these are obtained together by distillation and used mostly in the form of acetylated derivatives. Cedrol is used as such and, to a greater extent, as its acetate ester.



(4)

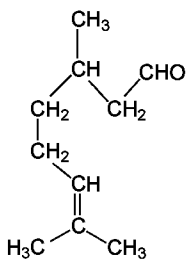


(5)

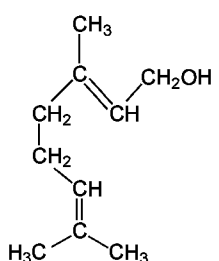


(6)

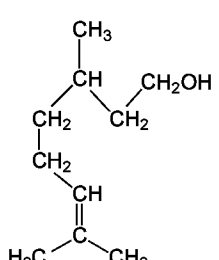
Citronella. Citronella oil is produced in Ceylon, China, Java, and Brazil by steam distillation of similar, but not identical, grasses. The main constituents of the oil are citronellal [106-23-0] (7), geraniol [624-15-7] (8), and citronellol [106-22-9] (9). The Javanese and, more recently, the Chinese oils have emerged as the most used materials in this family because they contain larger amounts of the desired aldehyde and alcohols. Citronellal, which should comprise 50–60% of a good quality oil, is mainly of the dextrorotatory form. It is desirable both as a perfume ingredient in its own right and for chemical conversion to hydroxycitronellal [107-75-5] (10), a long-lasting muguet odorant.



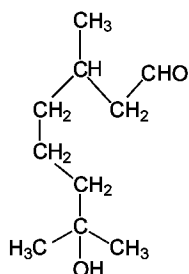
(7)



(8)

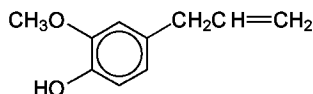


(9)

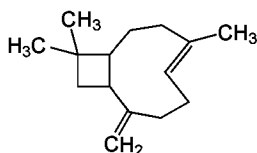


(10)

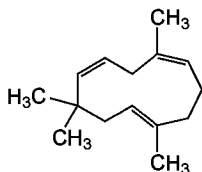
Clove Leaf Oil. Clove leaf oil is produced mainly in Madagascar and Indonesia. It is obtained by distillation of leaves and twigs of the *Eugenia caryophyllata*. The material from Madagascar is considered of superior quality to those from the other areas, because its eugenol [97-53-0] (**11**) content ranges between 82–92% ; in Indonesia, the eugenol content varies between 78–86% . Eugenol from this source is used as a chemical raw material for conversion to several derivatives, the most important of which is isoeugenol. The sesquiterpene section contains mainly caryophyllene [13877-93-5] (**12**), along with some humulene [6753-98-6] (**13**).



(11)



(12)

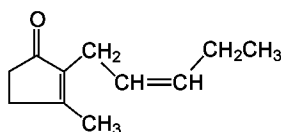


(13)

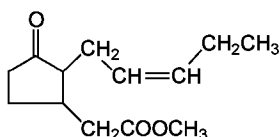
Galbanum. Galbanum gum is an exudate collected from large umbelliferous plants of the *Ferula* species, which grow wild in the Middle East. The gum is extracted with alcohol to produce a resinoid or steam-distilled to produce an oil. The odor of galbanum blends excellently with lilac fragrances. In modern perfumery, it is used to give a greenish top note. Galbanum oil and resinoid are complex mixtures from which many materials have been identified. Of these, a group of isomeric 1,3,5-undecatrienes has been identified as key odor contributors, in particular the 3(*E*),5(*Z*)-isomer [51447-08-6].

Geranium. Various perfume ingredients are produced from geranium, *Pelargonium graveolens*, in many parts of the world. These include a concrete from Morocco and an absolute produced from it. The most important geranium product by far is geranium bourbon, an oil produced by steam distillation of pelargonium leaves and branches. It originated from the island of Reunion in the Indian Ocean; however, most current production is from China and Egypt. In addition to its direct use in perfumes, geranium oil is fractionally distilled to provide, among other products, rhodinol. This material is comprised mainly of *L*-citronellol (**9**), although this by no means accounts for all of its fine odor quality. Rhodinol is highly desirable for fine perfumery applications, in particular as a base for rose, muguet, and other floral fragrances.

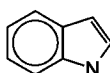
Jasmine. Jasmine is one of the most precious florals used in perfumery. The concrete of jasmine is produced by hydrocarbon extraction of flowers from *Jasminum officinale* (var. *Grandiflorum*). The concrete is then converted to absolute by alcoholic extraction. It is produced in many countries, the most important of which is India, followed by Egypt. Jasmine products are rather expensive and are produced in relatively small amounts compared with other materials. However, jasmine is particularly important in perfume creation for its great power and aesthetic qualities. Four of the principal odor contributors to jasmine are *cis*-jasnone [488-10-8] (**14**), methyl jasmonate [91905-97-4] (**15**), benzyl acetate [140-11-4], and indole [120-72-9] (**16**).



(14)



(15)

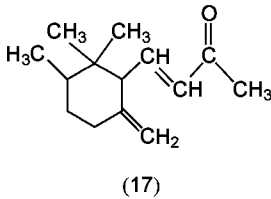


(16)

Lavandin. Lavandin, *Lavandula hybrida*, as a plant species is of recent origin, unknown until the late 1920s. It is a hybrid of two common lavenders, *Lavandula officinalis* and *Lavandula latifolia*. Lavandin is cultivated mainly in southern France and has become one of the most produced and used natural perfumery materials. The flowering tops of the shrub are used to produce a concrete, an absolute, and a steam-distilled oil; the last is by far the most used. Low cost and refreshing odor quality allow lavandin to be employed in a wide variety of perfume applications and at high concentrations. Chemically it is comprised of 30–32% linalool (3) and linalyl acetate (1), along with numerous other substances, mostly terpenic.

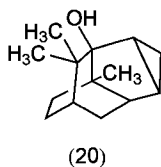
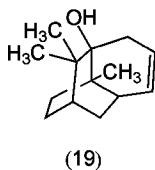
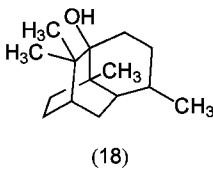
Oakmoss or Mousse de Chene. Oakmoss, *Evernia prunastri*, is a lichen that grows mostly on oak and spruce trees. It is collected mostly in the Czech Republic, Croatia, and Morocco. A cheaper quality, which is also called tree moss or mousse d'arbre, grows on spruce and pine trees. Oakmoss is worked into a variety of products, including a concrete, resinoid, and absolute, the last of which is the most used. Materials of this type are typically green in color and are thus limited in some functional perfumes. Decolorization techniques can be applied but some loss of odor quality occurs. Small amounts of oakmoss absolute are remarkably effective in perfumery for imparting a long-lasting, typical mossy note. They are used in the most expensive perfume products and blend well with other oriental or flowery notes. The main odor constituent of oakmoss is the ester methyl 2,4-hydroxy-3,6-dimethylbenzoate [115-10-6].

Orris. Orris is produced from rhizomes of *Iris pallida* and *Iris germanica*. The plants are found and cultivated mostly in Italy, but also in Morocco and China. It is used in perfumery as an absolute, a steam-distilled essential oil, and a concrete. The last material, which is a low melting solid (due to a high content of myristic acid) and therefore erroneously called a concrete, is by far the most used. Orris has a violet-like odor useful in fine perfumes, luxury soaps, and fragrances for powders and other cosmetic products. Its most important odor contributors are the irones, of which the most important isomer is [79-68-5] (17).



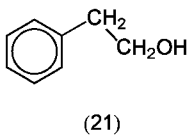
Orange Flower. Extraction of freshly picked flowers of the bitter orange tree, *Citrus aurantium* (subspecies *amara*), for the production of concrete is carried out mainly in Morocco and Tunisia. Most of this material is processed further to give orange flower absolute, one of the most important absolutes used in perfumes after rose and jasmine. It is highly valued in perfumery, even when used at low levels, for its long-lasting, rich, warm, yet delicate and fresh floralcy. The material is a complex mixture, to which methyl anthranilate [134-20-3], linalool (3), methyl jasmonate (15), and indole (16) are important odor contributors.

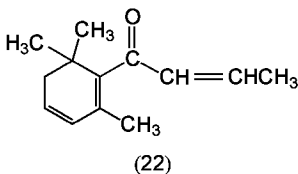
Patchouli. Patchouli oil is produced by steam distillation of the dry leaves of *Pogostemon cablin*, a shrub-like plant that originated in the Philippines and Indonesia. Most of the production in the 1990s is from Indonesia, although some is also produced in China. Patchouli oil has a wonderfully rich odor profile which is characterized as warm, sweet, herbaceous, spicy, woody, and balsamic. It is relatively inexpensive for a natural product and is usually available in abundance. For these reasons, patchouli oil is very widely used in many kinds of perfumes. Its main odor-donating constituents are a group of polycyclic alcohols. The best known of these, patchouli alcohol [5986-55-0] (18), is present in the oil to the extent of about 30% . However, it is believed that norpatchoulenol [41429-52-1] (19) and nortetrapatchoulol [62731-84-4] (20), which are present in smaller amounts (0.5–1%), are more important as odor contributors.



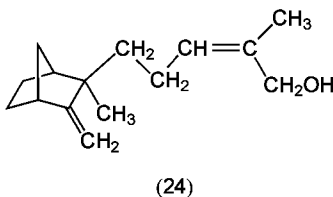
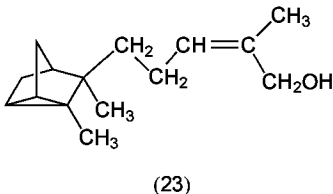
Petitgrain. Petitgrain oils are produced by steam distillation of leaves and twigs of the bitter orange tree, *Citrus aurantium*, the same species used to produce orange flower oil. The so-called biogarde oil is produced from the true bitter orange tree grown in southern France, Italy, Spain, and northern Africa. Petitgrain Paraguay, by far the most used material of this type, is produced from the bitter-sour variety in South America. The odor of these oils is fresh, bitter, and floral, with woody undertones. They are used widely in perfumery, particularly in citrus colognes and floral bouquet perfumes. In addition, the oils, mainly the Paraguay version, are redistilled to give a useful terpeneless oil. Important odor constituents of petitgrain oils are linalyl acetate (1), linalool (3), methyl anthranilate, geraniol (8), and nerol [106-25-2], 3,7-dimethyl-2,6-octadien-1-ol.

Rose. Rose is one of the most important florals in perfumery, the most valuable derivatives of which are produced from *Rosa damascena*, which is grown principally in Bulgaria, but also in Russia, Turkey, Syria, India, and Morocco. The concrete, absolute, and steam-distilled essential oil (rose otto) are particularly valuable perfume ingredients. Careful handling and processing of freshly picked flowers are required to produce these materials of warm, deeply floral, and rich odor quality. They are complex mixtures of which citronellol (9), geraniol (8), phenethyl alcohol [60-12-8] (21), and β -damascenone [23726-93-4] (22) (trace component) are important odor constituents.

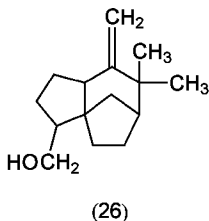
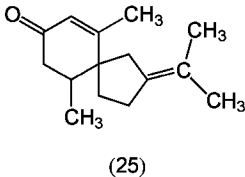




Sandalwood. Sandalwood is one of the oldest materials in fragrance use. Its oil is produced by steam distillation of coarsely ground wood and roots of *Santalum album*, a comparatively small, slow-growing tree. Its world production is concentrated in India and Indonesia, the latter being of less preferred quality. In order to obtain a good oil and high yield, only trees that are over 30 years in age are used. This limits the supply of the oil and makes it expensive. Although sandalwood oil is still valuable in some fine fragrance applications, much of the sandalwood odor of perfumes produced in the 1980s and 1990s is a result of the excellent synthetic materials that have been introduced. The main odor contributors to sandalwood oil are alpha-santalol [115-71-9] (23) and beta-santalol [77-42-9] (24).



Vetivert. Vetivert oil is steam-distilled from cleaned, dried, and chopped rootlets of *Vetiveria zizanioides*, a tall perennial grass normally grown for up to 20 months prior to harvesting. Most production of this material is in Haiti, Indonesia, Reunion, and, of a poorer quality, in China. The oil has a heavy woody-earthly odor and an undertone of precious wood. It is used as such, as distillation sections, and is treated with acetic anhydride to produce a mixture known as vetivert acetate. β -Vetivone [18444-79-6] (25) is probably the main odor contributor to this essential oil, along with a number of lower keyed vetiverols, a chemically interesting example of which is [16223-63-5] (26).



Violet Leaf. Violet leaf absolute is produced by the usual extraction methods from *Viola odorata* (var. *Victoria*). It is grown mainly in the south of France and Egypt. Although this material is not produced in large amounts, it is quite valuable in perfumery for its powerful green leafy and floral character, an odor that belongs to many floral bouquets. The principal odorant in violet leaf absolute is 2-*trans*-6-*cis*-nonadienal [557-48-2].

Ylang-Ylang. Flowers from the cultivated *Cananga odorata* tree, grown mostly in the Comoro islands and Madagascar, are the starting material for the concrete, absolute, and steam-distilled oil of ylang-ylang. The essential oil is the most important of these products. As with other florals, the flowers of ylang-ylang must be handled carefully and quickly to produce materials of high odor quality. The oil has a powerful floral and intensely sweet odor as well as some spicy and balsamic character. It is a valuable component of many different floral- and oriental-type perfumes. Benzyl acetate is the largest component (~30%) of the oil but is responsible for only a small part of its odor profile, which is the result of numerous minor constituents.

Aroma Chemicals. The use of aroma chemicals in perfumery has been growing since they were first introduced. A number of practical advantages account for this trend. Probably foremost among them is that the growing use of fragrance in the world outstripped the ability to produce enough natural materials, particularly the aesthetically important concretes and absolutes. Increasing world population and the need for food-farming have displaced some of the land area and labor required for production of perfume ingredients. Naturals, especially those produced from flowers, have become rather expensive due to limited supply and rising labor costs. Some essential oils, particularly those derived from slow-growing trees such as sandalwood, have also been limited in supply for environmental or political reasons. Quality control of synthetics is straightforward compared with naturals because the raw material and product compositions are much less complex. Also, synthetics are not subject to variation in quality and supply due to growing conditions.

One of the main reasons why aroma chemicals have grown to be such a large part of the fragrance industry is their availability. This has been possible because synthetic fragrance materials are produced from a wide variety of starting materials, from both petrochemical and renewable sources. The most important renewable source is turpentine, followed at some distance by cedarwood oil. The most produced synthetics as of 1995 reach worldwide production levels of several thousand tons per year, but most are produced in much lesser quantities. In general, batch-processing is used, although there are exceptions, usually for chemical reasons rather than economic ones. Thus, synthetic fragrance ingredients are specialty chemicals and as such do not place heavy demands on available feedstocks.

Aroma chemicals are not limited to any particular functional group, some being more common than others. The majority of materials contains a single oxygenated functional group, although there are also important materials containing two and even three. Most of these are manufactured by syntheses of one to four chemical steps; some are isolated by distillation from abundant natural sources, eg, cedarwood and clove leaf oils. Some hydrocarbons are important fragrance ingredients, as well as several sulfur or nitrogen-containing materials, but these are generally used in small amounts to provide special nuances. The molecular weight range is roughly from 100–300, and most of the materials have from 10–18 carbon atoms per molecule.

There are no simple rules for chemically defining a good odorant. Table 2 and Figure 1 present a number of widely used aroma chemicals by chemical type.

Table 2. Typical Aroma Chemicals by Functional Group

Name	CAS registry Number	Structure numbers ^a	Odor type
<i>Hydrocarbons</i>			
caryophyllene	[13877-93-5]	(12)	woody, spicy (cloves)
β-farnesene	[18794-84-8]	(27)	mild, sweet, warm
limonene	[138-86-3]	(28)	orange, citrus
α-pinene	[80-56-8]	(29)	piney, woody
β-pinene	[127-91-3]	(30)	
<i>Alcohols</i>			
Bacdanol	[29219-61-6]	(31)	sandalwood
citronellol	[106-22-9]	(9)	rosy, citrus
linalool	[78-70-6]	(3)	floral, citrus
phenethyl alcohol	[60-12-8]	(21)	floral, rosy
α-terpineol (R = H)	[98-55-5]	(32)	floral, lilac
<i>Aldehydes</i>			
2-methyl undecanal	[110-41-8]		fresh, citrus (orange)
citral	[5392-40-5]	(2)	citrus, lemon
hexyl cinnamic aldehyde	[101-86-0]	(33)	floral, jasmine
Isocyclocitral	[1423-46-7]	(34)	floral, carnation
Lilial	[80-54-6]	(35)	floral, muguet
10-undecenal	[112-45-8]		citrus, waxy
<i>Ketones</i>			
Cashmeran	[33704-61-9]	(36)	musky, sweet
α-ionone	[127-41-3]	(37)	floral, violet
Isocyclemone E	[54464-57-2]	(38)	amber, woody
Koavone	[86115-11-9]	(39)	woody, ambery, floral
muscone	[541-91-3]	(40)	musk
Tonalide	[21145-77-7, 1506-02-1]	(41)	musk
<i>Esters</i>			
benzyl acetate	[140-11-4]		floral, jasmine
cis-4- <i>t</i> -butylcyclohexyl acetate	[10411-92-4]		woody, floral
trans-4- <i>t</i> -butylcyclohexyl acetate	[1900-69-2]		
cedryl acetate	[77-54-3]	(42)	woody, cedar
Cyclacet	[5413-60-5, 2500-83-6]	(43)	green, woody
isobornyl acetate	[125-12-2]	(44)	pine needles
α-terpinyl acetate (R = acetyl)	[80-26-2]	(32)	herbaceous, piney
<i>Lactones</i>			
coumarin	[91-64-5]	(45)	sweet, hay
jasmine lactone	[34686-71-0]	(46)	floral, jasmine
muskalactone	[106-02-5]	(47)	musk
peach aldehyde	[104-67-6]	(48)	fruity, peach
<i>Ethers</i>			
Ambroxan	[3708-00-9]	(49)	amber, woody
Anther	[56011-02-0]	(50)	floral, hyacinth
Galaxolide	[1222-05-5]	(51)	musk
<i>Nitriles</i>			
cinnamonnitrile	[4360-47-8]	(52)	cinnamic, balsamic
geranonitrile	[5146-66-7]	^b	citrus, lemon
<i>Polyfunctionals</i>			
amyl salicylate	[2050-08-0]	(53)	floral, jasmine
isoeugenol	[97-54-1]	(54)	warm, spicy, floral
Hedione	[29852-02-6]	(55)	jasmin, lemon
heliotropine	[120-57-0]	(56)	sweet, floral
Lylal	[31906-04-4]	(57)	floral, muguet
vanillin	[121-33-5]	(58)	sweet, vanilla

^a See Fig. 1.
^b See structure (8); CN replaces CH₂OH.

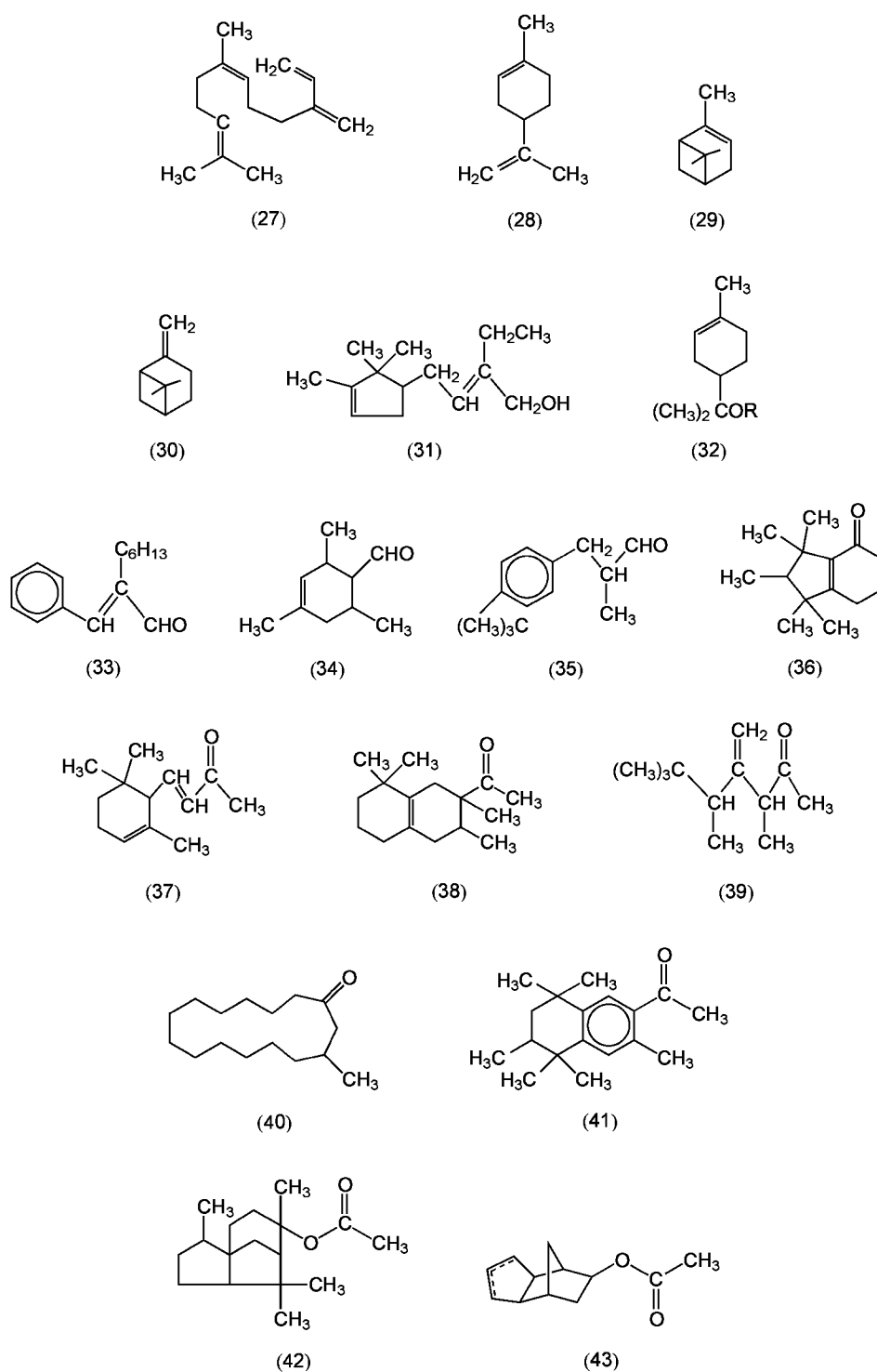
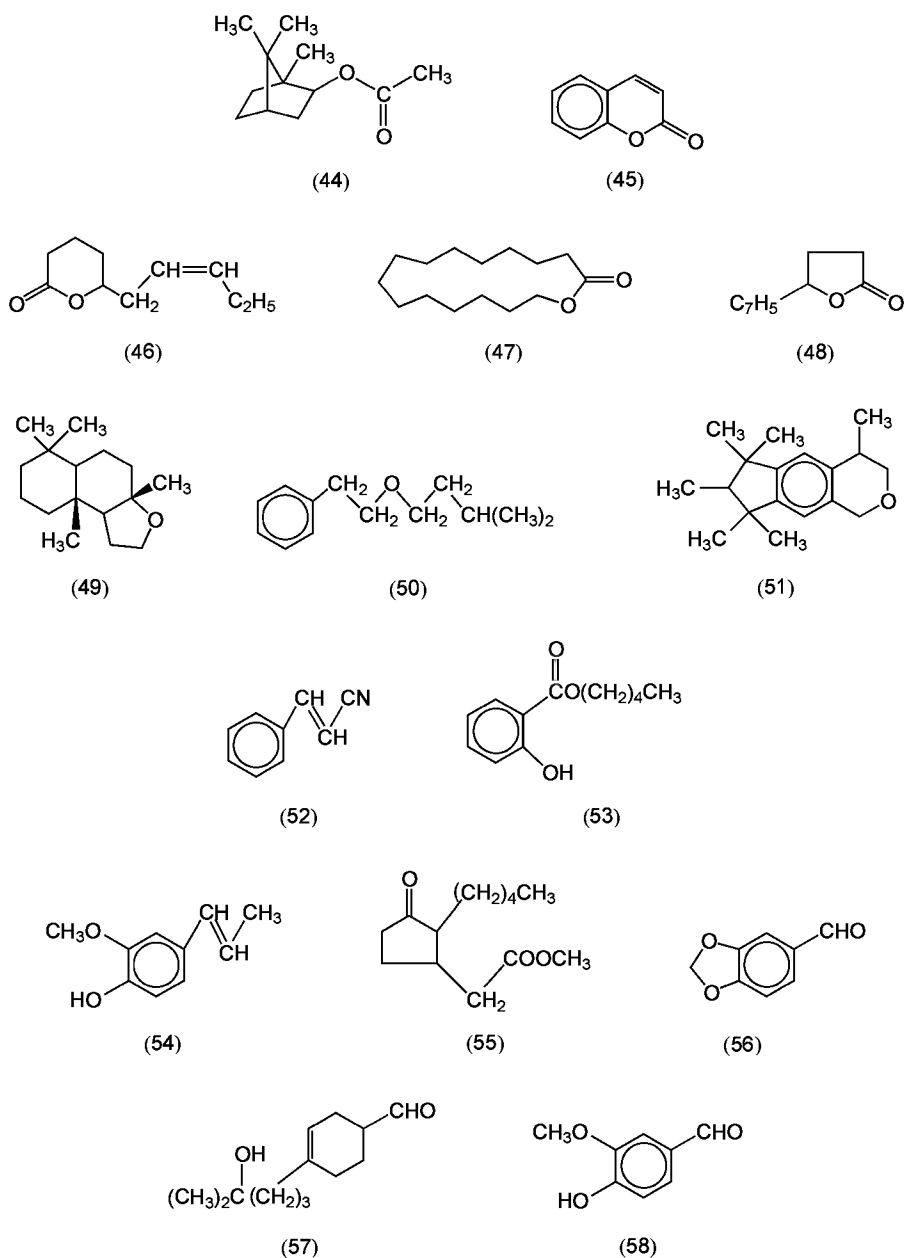


Fig. 1. Structures of some typical aroma chemicals. See Table 2.



Aroma chemicals are used in perfumes over a wide range of concentrations. The importance of a material to the overall creation should not be judged by the amount used. It is characteristic of chemical manufacturing in the fragrance industry that many materials are produced over a wide range of production volumes. This situation can change as larger manufacturers attempt to simplify their product lines by giving up some low volume products. Nevertheless, it is reasonable to assume that manufacturers of aroma chemicals will continue to produce a wide range of materials.

Manufacture and Quality Control

Perfumes are manufactured by blending ingredients as called for by the perfumers' formulas. Most ingredients are oily liquids; however, some are solids and therefore mixing must allow for dissolution and thorough blending. During this operation, some protection from air oxidation is advisable for safety reasons and to assure the quality of the finished product. Depending on the final application, perfumes can be produced in batches of several kilograms to several tons. Because individual ingredients are used over a wide range of concentrations, precise weighing of small and large quantities is required.

Quality Control. Reproducible production of perfumes requires careful quality control of all materials used as well as the compounding process itself. The use of analytical tools has increased over the years with their availability, but there can be no substitute for organoleptic evaluation. The human nose is far more sensitive than any analytical instrument for certain materials, yet it is also quite limited as a quantitative tool and is subject to fatigue. There are also well-documented examples of specific anosmias in individuals, ie, inability to smell certain odor types, which is somewhat analogous to color-blindness.

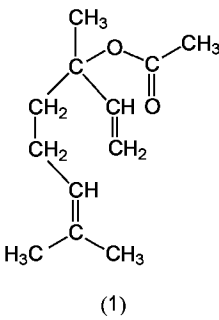
In a modern fragrance company, there can be several thousand ingredients either manufactured or purchased and kept in inventory for perfume compounding. Maintaining quality control by odor thus requires much organization, skill, and experience. Standard target samples of each ingredient and each finished fragrance must be kept on hand and properly stored for comparison with each new batch or shipment. Usually experienced evaluators perform the smelling and comparisons; various systems are used throughout the industry.

The techniques of analytical chemistry have been applied in the fragrance industry for as long as they have been available. Ingredients have long been characterized by wet chemical methods, color tests, distillation, and bulk analytical methods such as density and refractive index. These were as limited in value for fragrance materials as for other areas of organic chemistry. The rise of more specific instrumentation during the middle of the twentieth century brought great changes in the ability to test and standardize materials. Gas-liquid chromatography (glc) is particularly applicable to analysis of fragrances and fragrance materials. Refinements in the instrumentation of glc, especially capillary columns, have made welcome additions to quality-control laboratories. Nevertheless, odor quality cannot be ensured by even the best analytical techniques available in the 1990s. The fragrance industry therefore relies on both odor evaluation and analytical methods for control of ingredient and product quality.

Research

Analytical Chemistry. Research in the fragrance industry is rooted in the chemistry of natural products. Chemists have long analyzed essential oils and other fragrant materials derived from nature in order to determine their compositions and in particular to identify the odoriferous principles. This has produced a wealth of synthetic targets, many of which have become important aroma chemicals. Analysis of naturals has also been used to allow the preparation of synthetic reconstitutions or duplications, which can have great commercial importance if a material is in short supply or becomes very expensive. Duplications also allow important natural odor notes to be used in functional perfumery where supply or discoloration problems can arise.

There have been moves in the 1990s to replace natural materials of animal origin. This has occurred for humanitarian or economic reasons. One of the first and best known perfume ingredients to be eliminated from use was tincture of ambergris. The starting material for this was ambergris, a principal by-product of the whaling industry. Its use was therefore eliminated as part of international efforts to preserve whale populations. Tincture of ambergris has been replaced by formulas that include the most important contributors to ambergris odor, namely, α -ambrinol [41199-19-3] (59) and dihydro- γ -ionone [13720-12-2] (60). These are believed to form via oxidation and cyclization from the principal component, ambrein [473-03-0] (61) (1).



Most natural products of interest, especially some of the most valuable ones, are complex mixtures. Advances in the techniques of separation and identification have been quickly adopted and applied in order to examine the compositions of natural substances in ever greater depth (2). Gas-liquid chromatography is the essential tool for the investigation of fragrant materials because such materials must exhibit some degree of volatility in order to be perceived. The fused-silica capillary columns which have become available in the 1990s are a mainstay of analytical research. In order to obtain chemical structural information on the hundreds of materials that may be present in a complex mixture, gas chromatography is coupled with mass spectrometry (gc ms). Advances in computer methods for acquiring, sorting, and interpreting molecular fragmentation data from gc ms runs make it possible to perform thorough analyses with remarkable speed.

For those components that are not readily identified from their mass spectra alone, gas chromatography coupled with infrared spectroscopy (gc ir) has been applied. In these instruments, the infrared spectra are acquired in the vapor phase by Fourier transform methodology. Those components that resist identification by these methods must be isolated and analyzed by nuclear magnetic resonance (nmr). Instruments that utilize superconducting magnets and Fourier transform data acquisition, and which have small sample requirements and large resolving power, have allowed natural product chemists to identify many important and interesting materials.

Head space analysis techniques have also been used to investigate fragrant materials and have produced fascinating technical and commercial results. At first, such methods were used to investigate the top notes, ie, the most volatile part, of essential oils and other naturals. This was done in order to allow more complete reconstitution of naturals using synthetic ingredients. It was then found that these techniques could be applied directly to flowers and other parts of plants. This avoids the heating and other processing involved in producing natural extracts, so that duplications based on these results differ significantly from the older standard products. It also allows the duplication or imitation of flowers, herbs, fruits, etc, which do not produce satisfactory extracts either because they do not contain sufficient fragrant oil or because they produce undesirable odors during the extraction processes.

In an important next step, it has been found that flowers and other plant parts can be analyzed by using head space techniques without removing them from the living plant (3). It was immediately observed that there are remarkable differences in the volatile compositions observed from live and picked flowers. This is exemplified for jasmine flowers in Table 3. Reconstitutions produced from this information have provided perfumers with novel and fresh notes for use in their creations. This technique continues to be applied to many kinds and varieties of flowers, leaves (herbs, spices), and fruits. The reasons for the remarkable differences observed are not known.

Table 3. Head Space Constituents of Jasmine Flowers

Chemical component	Living flower, %	Picked flower, %
6-methyl-5-hepten-2-one	0.2	
cis-3-hexenyl acetate	0.2	
benzyl alcohol		4.0
ocimene (cis and trans)	0.2	1.1
benzyl acetate	60.0	40.0
linalool	3.0	30.0
indole	11.0	2.0
cis-jasmone	3.0	
3,5-dimethyl-2-ethylpyrazine		0.5
methyl jasmonate	0.3	

Synthesis. Exploratory research has produced a wide variety of odorants based on natural structures, chemicals analogous to naturals, and synthetic materials derived from available raw materials and economical processing. As in most areas of the chemical industry, the search for new and useful substances is made difficult by the many materials that have been patented and successfully commercialized (4). In the search for new aroma chemicals, many new materials are prepared for screening each year. Chemists who perform this work are involved in a creative exercise that takes its direction from the commercial sector in terms of desirable odor types and specific performance needs. Because of economic limitations, considerations of raw material costs and available processing methods may play a role early in the exploratory work.

Initial evaluations of chemicals produced for screening are performed by smelling them from paper blotters. However, more information is necessary given the time and expense required to commercialize a new chemical. No matter how pleasant or desirable a potential odorant appears to be, its performance must be studied and compared with available ingredients in experimental fragrances. A material may fail to live up to the promise of its initial odor evaluation for a number of reasons. It is not at all uncommon to have a chemical disappear in a formulation or skew the overall odor in an undesirable way. Some materials are found to be hard to work with in that their odors stick out and cannot be blended well. Because perfumery is an individualistic art, it is important to have more than one perfumer work with a material of interest and to have it tried in several different fragrance types. Aroma chemicals must be stable in use if their desirable odor properties are to reach the consumer. Therefore, testing in functional product applications is an important part of the evaluation process. Other properties that can be important for new aroma chemicals are substantivity on skin and cloth, and the ability to mask certain malodors.

Structure-Odor Correlations and Olfactory Receptors. The issue of structure-odor correlation is one that continues to fascinate and frustrate fragrance chemists. In certain odor areas, particularly for musk (5) and amber (1) odors, much work has been done and some correlations have been made. In these areas, chemical structures are rather well-defined and many are rigid in nature, thus simplifying conformational questions. Much careful synthetic work has been done in conjunction with some of these studies. It has been shown that odor strength and character can be exquisitely sensitive to very small structural changes in some, but not all, cases. This applies to stereochemical and chiral differences (6) as well as other simple changes such as adding, removing, or changing the position of a methyl group. Computer methods have been applied by using such techniques as molecular modeling, pattern recognition, and molecular orbital calculations in attempts to correlate various molecular properties with odor quality. These efforts cannot be

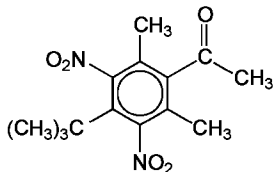
generalized over a range of odor types and are quite limited in predictive value.

The sense of smell is a chemical sense in that the molecules of an odorant must come in contact with some sorts of receptors, presumably located on the outside of neurons in the olfactory epithelium. The pathway of neuronal signals to the brain via the olfactory bulb has been known for some time, but until the early 1990s little was understood about the mechanisms of odor recognition and signal transduction. However, by using the techniques of modern biochemistry, important findings have been made and the beginnings of important insights are emerging (7,8).

Olfactory receptors have been a subject of great interest (9). Much that has been postulated was done by analogy to the sense of sight in which there are a limited number of receptor types and, as a consequence, only three primary colors. Thus attempts have been made to recognize primary odors that can combine to produce all of the odors that can be perceived. Evidence for this includes rough correlations of odors with chemical structural types and the existence in some individuals having specific anosmias. Cross-adaptation studies, in which exposure to one odorant temporarily reduces the perception of a chemically related one, also fit into this hypothetical framework. Implicit in this theory is the idea that there is a small number of well-defined odor receptors, so that eventually the shape and charge distribution of a specific receptor can be learned and the kinds of molecular structures for a specified odor can be deduced.

There appear to be several hundred, perhaps more than one thousand, olfactory receptors (10) which presents a new way to consider the sense of smell. This result is consistent with the fact that smell is a primitive sense, acute in lower animals that have small brains. Among the ramifications of this discovery is the knowledge that much of the information needed to differentiate odors is obtained at the point of contact between odorant molecules and receptor neurons. Considering that there are so many receptors, it is reasonable to assume that the signal leaving the nose contains most of the information needed by the brain to recognize an odor. It can be seen from Table 2 and Figure 1 that most odorants used in perfumes are rather nonpolar and therefore likely to have weak, hydrophobic, and fleeting interactions with receptor proteins.

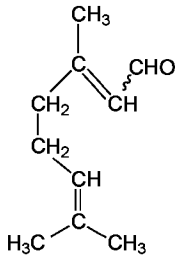
Thus it appears that perceived odors can result from a combination of many receptor interactions, which give rise to complex neuronal signals. Under this scenario, it is not at all clear if the characterization of a single receptor or several receptors would be helpful in predicting the odor caused by a given chemical entity. Many structure–odor researchers in the past have looked for similarities in the structures of rather disparate chemicals having similar odors, for example the musk odorants muscone [541-91-3] (40), Galaxolide [1222-05-5] (51), and musk ketone [81-14-1] (62). The new theory suggests that there may be more than one combination of receptors that can produce a given type of odor perception, or more than one way to stimulate a given combination of receptors. Chemists working in this area have begun to take these new findings into account (11). It remains to be seen what impact this will have on fragrance chemical research.



(62)

Process Research and Development. In the fragrance industry, as in other parts of the chemical industry, chemical processing costs must be consistent with values of the materials produced. It is a common fallacy, probably inferred from the costs of couture perfumes, that fragrance ingredients are expensive. This is true for certain natural materials produced laboriously from flowers, but is certainly not the case for the vast majority of aroma chemicals in use. The costs of most large-volume (roughly >50 t yr) aroma chemicals are quite competitive. Consequently, the fragrance industry has invested much effort in process research and development. In addition to looking for ways to improve yields and throughputs, significant changes have taken place in the scale of manufacture, the equipment used, and the manufacturing operations themselves. As these changes occur, maintaining product quality has been a critical concern.

In several important cases, new synthetic strategies have been developed into new production schemes. An outstanding example of this is the production of an entire family of terpene derivatives from α -pinene (29), the major component of most turpentines, via linalool (3) (12). Many of these materials had been produced from β -pinene, a lesser component of turpentine, via pyrolysis to myrcene and further chemical processing. The newer method offers greater manufacturing flexibility and better economics, and is environmentally friendly in that catalytic air oxidation is used to introduce functionality.



(2)

In addition to large-scale process work, there is also some effort expended in providing synthetic methods for producing small amounts (<100 kg yr) of materials needed for replacement of naturals that are being discontinued or in short supply. Examples are ambrinol (59) and dihydro- γ -ionone (60) for the replacement of ambergris. More recently, tincture of tonquin musk has been replaced by using, among other materials, mixtures of macrocyclic ketones and alcohols which have been found in the natural substance.

Most aroma chemicals are relatively high boiling (80–160°C at 0.4 kPa $\approx$ 3 mm Hg) liquids and therefore are subject to purification by vacuum distillation. Because small amounts of decomposition may lead to unacceptable odor contamination, thermal stability of products and by-products is an issue. Important advances have been made in distillation techniques and equipment to allow routine production of 5000 kg or larger batches of various products. In order to make optimal use of equipment and to standardize conditions for distillations and reactions, computer control has been instituted. This is particularly well suited to the multipurpose batch operations encountered in most aroma chemical plants. In some instances, on-line analytical capability is being developed to work in conjunction with computer controls.

Physiological and Psychological Effects of Fragrance. The sense of smell is much more important for the preservation of life for most animals than it is for human beings. It is critical, among other things, to feeding, safety, recognition of individuals, and reproduction. As humans evolved, the importance and acuity of the sense of smell decreased, though it still can have powerful effects on humans, such as the stimulation of memory. Considerable research effort has been and continues to be made to evaluate the role of fragrance in human behavior.

Some efforts to determine physiological effects have focused on peripheral measurements such as pulse rate, blood pressure, and galvanic skin response. Pleasant odors have little or no effect on these functions, whereas unpleasant odors can produce alarm reactions. Brain waves, measured by electroencephalography (eeg), are one area where physiological changes have been observed upon exposure to fragrances (13). Odors have been shown to evoke event-related potentials (ERPs) in the brain. These vary with the pleasantness of the odor and whether or not an odor is also a trigeminal stimulus. Odors also affect the brain's spontaneous eeg and ERPs to visual stimuli, which indicates that they have the potential to influence the brain's processing of

information. This line of research is quite new, however, and much more work remains before brain mechanisms of olfaction can be addressed.

Faced with the limits of physiological measurements in determining the effects of fragrance on humans, researchers have turned to psychological measurements and produced some interesting results (14). Such methods as monitoring behavior and various types of self-reporting (eg, questionnaire) before and after exposure to many odor types have been used; requiring subjects to relate odors to various kinds of images has also been fruitful. The results indicate that odors have small priming effects on mood and can affect behavior in predictable ways. Pleasant odors tend to improve mood. For example, several studies have shown that pleasant odors, compared to unpleasant ones, elicit more happy memories, enhance creative performance, and result in more positive evaluations (photographs or computer images) of people. Social psychological studies have examined the effects of odors on simulated negotiations; subjects exposed to pleasant odors were more cooperative and less prone to confrontational approaches. Thus, by studying and quantifying the effects of fragrance and fragrance ingredients on mood, it may be possible to add a new dimension to the performance of perfumes.

Economic and Market Information

Attempts to estimate worldwide merchant sales of fragrances are complicated by various factors. All principal suppliers of fragrances are also engaged in flavor business and most are also producers, to differing degrees, of aroma chemicals and naturals (used for both flavors and fragrances) which are used internally and sold throughout the industry (see FLAVORS AND SPICES). The industry has enjoyed fairly rapid growth rates, but figures from year to year can be highly distorted by relative currency changes. For purposes of this discussion, only the value to the fragrance supplier is considered. Retail sales of fine perfumes would give substantially higher numbers. For 1990, worldwide sales of compounded fragrances have been estimated at \$2800 million and sales of aroma chemicals at \$1500 million (15). It is a reasonable extrapolation that these numbers were about 20% higher in 1994. In 1990, the value of essential oils and other naturals sold was \$1350 million. A breakdown of product types where perfumes are used as determined for 1987 is as follows: women's and men's fragrances (perfumes, colognes, etc), at 26% of the market share; cosmetics and toiletries, at 26%; soaps (including toilet soaps), as well as laundry and dishwashing products, at 34%; and cleaning, disinfecting, and polishing products, air fresheners, and industrial products, at 14% (15).

Safety, Regulatory, and Environmental Aspects of the Industry

The fragrance industry has a long record of safety, largely on account of the nature and sources of its ingredients, and how its products are used. The approach to product safety, used successfully for many years, is based on individual ingredients rather than finished perfumes. Far more testing of fragrance formulations would be required than is done for ingredients because the rate of new fragrance creation is high. By examining individual ingredients and setting appropriate limits on their use, it is possible to ensure the safety of fragrances as they are created. This approach is accepted by relevant governmental agencies around the world, such as the U.S. Food and Drug Administration.

The industry supports two key organizations that strengthen scientific criteria and develop guidelines for safe and environmentally sound use of fragrances. The Research Institute for Fragrance Materials (RIFM) is an internationally recognized scientific organization that collects, generates, and disseminates information on the safety of perfume ingredients. This information may originate from published or unpublished sources, or through RIFM's ongoing research program. The findings are reviewed by an expert panel of academicians and published in peer-reviewed journals such as *Food & Chemical Toxicology*. The activities of RIFM are harmonized with those of the International Fragrance Association (IFRA). The IFRA, whose members come from various national associations of fragrance manufacturers (eg, from the Netherlands, France, Germany, United States, and Japan), is concerned with all aspects of safety evaluation and regulation in the industry, into which it has introduced self-regulatory discipline. Its primary function is the formulation and continuous updating of the Code of Practice for the Fragrance Industry. The Code provides guidelines on manufacturing practices as well as safe usage, toxicological methodology, and safety assessments for fragrance materials. These guidelines have been followed by the fragrance industry since the early 1970s.

Fragrances must comply with all applicable regulations and legislation that address occupational and consumer health, safety, and environmental concerns. Many countries have adopted chemical substance inventories in order to monitor use and evaluate exposure potential and consequences. For most fragrance applications, all ingredients must be on these lists. New substances must be subjected to premanufacturing or premarketing notification (PMN). PMN requirements vary by country and the predicted volume of production; they require assessment of environmental and human health-related properties, and reporting the results to designated governmental authorities.

Perfumes are also impacted by legislation that regulates specific products which may contain fragrances; such legislation includes the U.S. Food, Drug and Cosmetic Act and the European Community Cosmetic Directive. Under the latter regulation, a list of fragrance ingredients is being added to the European Community Cosmetic Ingredient Inventory to assist regulators and health officials in evaluating the safety of such products. An example of an environmental issue that may affect fragrance use and that is in legislation is the reduction of the atmospheric release of volatile organic chemicals (VOCs) from consumer products and other sources. Because fragrances are recognized as unique, essential components of consumer products, and generally used at low levels, they are given specific exemptions from most VOC regulations. Furthermore, the amount of VOCs released from consumer products is far less than those from other synthetic and natural sources.

Evaluation of the safe use of perfumes is an ongoing process. It is conducted mainly through self-disciplinary efforts of the fragrance industry and involves continuing investigation of the safety of new and existing ingredients. The manufacture and use of perfumes must comply with the growing body of environmental and human health-related regulations worldwide.

BIBLIOGRAPHY

"Perfumes" in *ECT* 1st ed., Vol. 10, pp. 7–17, by P. Chaleyer, Ph. Chaleyer, Inc.; "Perfumes (Synthetics and Isolates)" in *ECT* 1st ed., Vol. 10, pp. 17–38, by P. Z. Bedoukian, Faberge, Inc.; "Perfumes" in *ECT* 2nd ed., Vol. 14, pp. 717–746, by E. Shiftan, International Flavors & Fragrances, Inc.; in *ECT* 3rd ed., Vol. 16, pp. 947–971, by W. I. Taylor, B. Chant, and G. van Loveren, International Flavors & Fragrances, Inc.

1. G. Olhoff, in E. T. Theimer, ed., *Fragrance Chemistry: The Science of the Sense of Smell*, Academic Press, Inc., San Diego, Calif., 1982, p. 535.
2. J. P. Walradt, in Ref. 1, p. 575.
3. B. D. Mookherjee, R. W. Trenkle, and R. A. Wilson, *Pure Appl. Chem.* **62**, 1357 (1990).
4. P. Z. Bedoukian, *Perfumer Flavor.* **18**, 1 (1993) and prior articles in the series.
5. G. Klopman and D. Ptchelintsev, *J. Agric. Food Chem.* **40**, 2244 (1992).
6. M. H. Bolens, in Ref. 4.
7. R. R. H. Anholt, in M. J. Serby and K. L. Chobor, eds., *Science of Olfaction*, Springer-Verlag, New York, 1992, p. 51; W. T. Nickell and M. T. Shipley, *Ibid.*, p. 172.
8. R. R. Reed, in R. Hopp and K. Mori, eds., *Recent Developments in Flavor and Fragrance Chemistry*, VCH Publishers, New York, 1993, p. 275.
9. J. A. Amooore, in Ref. 1, p. 28.
10. L. Buck and R. Axel, *Cell* **65**, 175 (1991).
11. R. Pelzer and co-workers, in Ref. 8, p. 29.
12. *Chem. Eng. News*, 5 (Nov. 22, 1982).
13. T. S. Lorig, in S. Van Toller and G. H. Dodd, eds., *Fragrance: The Psychology and Biology of Perfume*, Elsevier Science Publishing Co., Inc., New York, 1992, p. 161; G. Kobal and T. Hummel, *Ibid.*, p. 175; S. Van Toller, S. Hotson, and M. Kendal-Reed, *Ibid.*, p. 195.
14. H. Ehdichman and L. Bastone, in Ref. 7.
15. L. Unger, *Perfumer Flavor.* **14**, 42 (1989).

General References

S. Arctander, *Perfume and Flavor Materials of Natural Origin*, M. Arctander, Elizabeth, N.J., 1960.
S. Arctander, *Perfume and Flavor Chemicals*, Vols. 1 and 2, M. Arctander, Elizabeth, N.J., 1969.
K. Bauer and D. Garbe, *Common Fragrance and Flavor Materials*, VCH Verlagsgesellschaft, Weinheim, Germany, 1985.
P. Z. Bedoukian, *Perfumery and Flavoring Synthetics*, Allured Publishing Corp., Wheaton, Ill., 1986.
M. Billot and F. V. Wells, *Perfumery Technology: Art, Science, Industry*, Ellis Horwood Ltd., Chichester, U.K., 1975.
W. A. Poucher, *Perfumes, Cosmetics and Soaps*, 9th ed., Vols. 1–3, Chapman and Hall, London, 1992.
R. R. Calkin and J. S. Jellinek, *Perfumery Practice and Principles*, John Wiley & Sons, Inc., New York, 1994.

William L. Schreiber
International Flavors & Fragrances, Inc.



PERIODIC ACID, PERIODATES.

See IODINE AND IODINE COMPOUNDS.

PERMANGANATES.

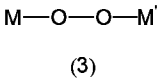
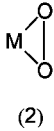
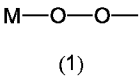
See MANGANESE COMPOUNDS.

PEROXIDES AND PEROXIDE COMPOUNDS

Inorganic peroxides,
Organic peroxides,

INORGANIC PEROXIDES

A peroxide or peroxy compound contains at least one pair of oxygen atoms, bound by a single covalent bond, in which each oxygen atom has an oxidation number of -1 . The peroxide group can be attached to a metal, M, through one (1) or two (2) oxygen atoms, or it can bridge two metals (3):



Peroxides should be distinguished from several other types of compounds having similar names. The higher oxides of lead, manganese, and other elements, although sometimes called peroxides, are not peroxides as defined herein because these contain no oxygen–oxygen bond. Similarly, compounds such as the perchlorates and permanganates are not peroxides. Rather, in these cases, the prefix *per* denotes the fully oxidized state for the central atom. It is preferable for true peroxides to be designated by the prefixes *peroxo* or *peroxy*. In the IUPAC nomenclature, *peroxo* is used for inorganic compounds, *peroxy* for organic compounds.

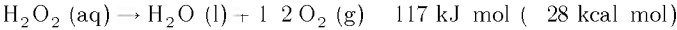
All the simple peroxides form hydrogen peroxide (qv) on contact with water. Some, eg, sodium peroxide, hydrolyze almost instantaneously:



Others, eg, the peroxodisulfates, hydrolyze slowly at room temperature:



Many inorganic peroxides tend, as does H_2O_2 , to decompose evolving oxygen:



Whereas decomposition is slow in pure solutions, it is accelerated enormously by alkali and traces of many metal ions. Indeed, hydrolysis to H_2O_2 , followed by its disproportionation, is the main path for decomposition of inorganic peroxides.

In hydrogen peroxide, the oxygen–oxygen bond strength is 209 kJ/mol (50 kcal/mol), which is approximately half of a normal bond strength for a single covalent oxygen bond. The oxidizing power of the peroxides results from this low bond energy as well as the high energies of O–M, O–C, and O–H bonds.

All of the commercial inorganic peroxy compounds except hydrogen peroxide are described herein, as are those commercial organic oxidation reactions that are believed to proceed via inorganic peroxy intermediates. Ozonides and superoxides are also included, but not the dioxygen complexes of the transition metals.

Group 1 (1A) Peroxides

Peroxides of all the alkali metals having the formula M_2O_2 are known. There are several general methods of preparation: reaction of the metal and oxygen, reaction of the metal monoxide and oxygen, thermal decomposition of the superoxide, and reaction of alkaline solutions of the metal and hydrogen peroxide. The last method generally gives peroxohydrates and/or hydrates of the peroxides, from which the hydrogen peroxide or water can be removed by mild heat treatment or vacuum. There is also an organic route to sodium peroxide octahydrate. A general account of all these peroxides is available (1).

Alkali metal peroxides are stable under ambient conditions in the absence of water. They dissolve vigorously in water, forming hydrogen peroxide and the metal hydroxide. They are strong oxidizing agents and can react violently with organic substances. Only lithium peroxide and sodium peroxide have been commercialized.

Lithium Peroxide. Lithium peroxide [12031-80-0], Li_2O_2 , is used in space technology because it absorbs carbon dioxide and liberates oxygen (see OXYGEN GENERATION SYSTEMS). This peroxide, also used for hardening certain plastics, is a white or pale yellow solid, stable at ambient temperature, and not hygroscopic. On heating to about 300°C, it loses oxygen and forms lithium monoxide. The commercial product contains about 96% Li_2O_2 . Unlike the other alkali peroxides, the lithium salt cannot be made by direct reaction between the metal and oxygen. It is made commercially by reaction of aqueous lithium hydroxide and hydrogen peroxide, which yields the peroxohydrate trihydrate, and is then dehydrated.

Lithium peroxide is a strong oxidizer and can promote combustion when in contact with combustible materials. It is a powerful irritant to skin, eyes, and mucous membranes (2); protective clothing should be worn when handling lithium peroxide. The LD_{50} has not been determined, and there is no designated threshold limit value (TLV). However, 5 g of many lithium compounds can be fatal.

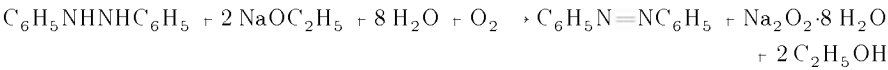
Commercial lithium peroxide has been assigned UN No. 1472 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances. It is manufactured by Chemetall AG (Germany) and Lithium Corp. of America (United States). The U.K. price in 1994 was 148–198 kg (\$70–285 kg), depending on quantity.

Sodium Peroxide. Sodium peroxide [1313-60-6], Na_2O_2 , is a pale yellow solid, stable at ambient temperature, and hygroscopic. On heating, it starts to liberate oxygen at about 300°C and decomposes rapidly above its melting point of 460°C. It dissolves rapidly in water, forming a solution of hydrogen peroxide and sodium hydroxide. The hydrogen peroxide then decomposes, liberating oxygen. When dissolving in water, the peroxide should always be added to the water, not vice versa, because of the large amount of heat evolved.

The commercial product is a powder containing a minimum of 96% Na_2O_2 and approximately 20% active oxygen. It is made commercially by oxidizing the molten metal with either oxygen or air enriched in oxygen. Early industrial history (1) and manufacturing details (3) are available. Sodium

peroxide used to be an important industrial chemical. It was used in bleaching wood pulp and linen that require strong alkalinity, as well as for the manufacture of sodium peroxoborates by the Acid and Duplex processes (4) and of sodium carbonate peroxohydrate. In the 1960s, with the advent of relatively inexpensive hydrogen peroxide from the anthraquinone process, sodium peroxide became uneconomic. As of the mid-1990s, it has only a few special applications, including chemical analysis and the extraction of platinum from its ores by the Leidie process.

The octahydrate of sodium peroxide [12136-94-6], Na₂O₂·8H₂O, was made commercially in Finland in the 1950s (5) by reaction of hydrazobenzene and sodium ethoxide:



This octahydrate, and the other hydrates and peroxohydrates of sodium peroxide that are known, no longer have commercial significance.

Although neither inflammable nor self-igniting, sodium peroxide is highly inflammable when mixed with oxidizable substances. Such mixtures burn violently, even in the absence of air. True sodium peroxocarbonates can be formed under the influence of atmospheric moisture and carbon dioxide. At temperatures >50° C and when exposed to pressure or friction, these peroxocarbonates can decompose and generate flame.

Sodium peroxide is a powerful irritant to skin, eyes, and mucous membranes (2); protective clothing should be worn when handling. This product has been assigned UN No. 1504 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances. It is manufactured by Metaux Speciaux (France), Nippon Soda (Japan), and Tosoh Corp. (Japan). Trade names that have been used for this product are Solozone and Flocool 180.

Group 2 (IIA) Peroxides

All the elements of Group 2 form peroxides, with the exceptions of beryllium and radium. There are two general methods of preparation: reaction of the metal or monoxide with oxygen, and reaction of the hydroxide with aqueous hydrogen peroxide. These peroxides are more stable in the presence of water than the Group 1 peroxides, primarily because of insolubility in water. Calcium peroxide is used on a large scale; magnesium, strontium, barium, and zinc peroxides have small-scale uses; whereas cadmium and mercury peroxides have no commercial uses at all. A general account of these peroxides is available (1).

In spite of widespread usage of these compounds, the structures of only the calcium, barium, and strontium compounds are reasonably well-established. The materials are generally made by triturating the oxides, or hydroxides, with aqueous hydrogen peroxide and drying the solid products. The commercial products are typically mixtures of the peroxides with varying amounts of hydroxides, oxides, carbonates, hydrates, and peroxohydrates.

Magnesium Peroxide. Magnesium peroxide [1335-26-8] and [14452-57-4], MgO₂, used in medicine as a stomach antacid and as an antiseptic (see DISINFECTANTS AND ANTISEPTICS), has not been prepared in the pure state. Commercial magnesium peroxide is made by mixing a light grade of magnesium oxide with hydrogen peroxide and drying the slurry. The product is a white powder containing about 25° o MgO₂ and 7° o active oxygen. The remaining constituents are magnesium oxide, magnesium hydroxide, and magnesium carbonate. This material is sparingly soluble in water but reacts with water slowly, forming hydrogen peroxide and liberating oxygen gas.

There are minor uses for magnesium peroxide in household products, veterinary medicine, and metallurgy (qv). Magnesium peroxide is a strong oxidizer and can cause fire when in contact with combustible materials. It is a powerful irritant to skin, eyes, and mucous membranes (2); protective clothing should be worn when handling. It has been assigned UN No. 1516 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances. Magnesium peroxide is made by Solvay Deutschland (Germany), which uses the trade name IXP^{ER} 25 M, and by L'Air Liquide (France). The price of the nominally 30° o-grade in France in 1994 was Fr. 120 kg (\$14 kg).

Calcium Peroxide. Pure calcium peroxide [1305-79-9], CaO₂, has been prepared, but the commercial product is a mixture made by reaction of calcium hydroxide and hydrogen peroxide. Commercial material contains either 60 or 75° o CaO₂; the remainder is a poorly defined mixture of calcium oxide, hydroxide, and carbonate. A well-defined octahydrate [60762-59-6], CaO₂·8H₂O, can be crystallized from aqueous systems.

An important application of calcium peroxide is for curing the polysulfide sealants (qv) used in double-glazing window units. Calcium peroxide is also used at several gold mines in Australia to increase the recovery of gold and reduce the consumption of cyanide. Calcium peroxide is preferred over hydrogen peroxide because the former causes less oxidation of the cyanide. Calcium peroxide is made *in situ* by mixing a slurry of calcium hydroxide and aqueous hydrogen peroxide immediately before use (6). Solid calcium peroxide can also be used in the heap-leaching of lean gold ores. A proprietary form of calcium peroxide for this purpose is sold by FMC (United States) under the trademark PermeOx.

PermeOx is also used to improve the bioremediation of soils contaminated with creosote or kerosene (see BIOREMEDIATION (SUPPLEMENT)), to deodorize sewage sludges and wastewater (see ODOR MODIFICATION), and to dechlorinate wastewater and effluents. A special formulation of calcium peroxide, made by FMC and sold in the United States under the trademark Trapzene, is used for removing metal ions from acidic waste streams such as coal ash leachate and acid mine drainage (see WASTES, INDUSTRIAL).

Calcium peroxide has several horticultural and agricultural applications, particularly in Japan. Usually used in the form of granules, it acts by providing extra oxygen for germinating plants and other organisms. In Japan, it is used in the cultivation of tomatoes, cucumbers, tobacco, and prawns; in Europe, it has been used as a coating on beet seeds. A considerable amount of developmental work has been done on other agricultural applications, including composting and the cultivation of rice and trees; however, as of this writing (ca 1995), these applications have not been commercialized.

Calcium peroxide has been used for many years as a dough conditioner in the United States (7), but not in Europe, where this use is not permitted. The application was first mentioned in 1921 (8) and described in 1930 (9). In dough, the calcium peroxide acts primarily by converting the disulfide bonds of gluten into sulphydryl groups, thereby preserving an open dough structure which retains more moisture than untreated dough. Other advantages are also provided in bread manufacture. A food-grade of calcium peroxide for bakery use is made by FMC in the United States (see BAKERY PROCESSES AND LEAVENING AGENTS; FOOD ADDITIVES). Another industrial application of calcium peroxide is as an oxidizing agent in the production of certain titanium–aluminum alloys.

Calcium peroxide is among the safest of the inorganic peroxides, presenting no significant hazard with regard to skin contact or absorption, inhalation, and ingestion; but it may be irritating to the skin under humid conditions. Airborne dust is irritating to the eyes, nose, throat, and lungs, but poses no significant long-term inhalation hazard. It should be handled only in well-ventilated areas and dust-controlled to below the TLV of 5 mg m⁻³; goggles, dust mask, and gloves should be worn when handling. This product has been assigned UN No. 1457 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances.

Calcium peroxide is made by FMC, L'Air Liquide, Nippon Peroxide (Japan), Shimakyo Chemical (Japan), Solvay Deutschland, Solvay Interox (United Kingdom), and Tomita Seiyaku (Japan). Trade names used for calcium peroxide are Calper, IXP^{ER} 60 C, IXP^{ER} 75 C, PermeOx, and Trapzene. The world market for calcium peroxide in 1992 was about 2000 t. The U.S. prices in 1994 were, for the food-grade (75° o CaO₂), \$3.97 kg; and for the technical grade (60° o), \$2.97–3.30 kg.

Strontium Peroxide. Commercial strontium peroxide contains about 85° o SrO₂ and 10° o active oxygen. It can be made by heating strontium oxide in the presence of oxygen gas under 20 MPa (200 atm) pressure, or by reacting a soluble strontium salt with hydrogen peroxide. The only substantial application for this compound is in pyrotechnics (qv). Strontium peroxide [1314-18-7] produces a red color in flames.

Strontium peroxide is a strong oxidizer and can cause fire when in contact with combustible materials. It is a powerful irritant to skin, eyes, and mucous membranes (2); protective clothing should be worn during handling. It has been assigned UN No. 1509 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances. It is made by Solvay Deutschland, which uses the trade name IXP^{ER} 85 S, and by L'Air Liquide. The price in France in 1994 was Fr. 130 kg (\$15 kg).

Barium Peroxide. Barium peroxide [1304-29-6], BaO₂, was an important compound in the late nineteenth and early twentieth centuries

because it was used in the manufacture of oxygen by the Brin process and of hydrogen peroxide by the Barium process (4). It is the only Group 2 peroxide that can be made by heating the monoxide in air. The monoxide is converted to the peroxide at 500–550°C; further heating to ca 700°C reverses the reaction. Small concentrations of steam can increase the rate of peroxide production.

The commercial product is a dull yellow powder containing about 90% BaO₂ and about 8.5% active oxygen; the remainder is mainly barium carbonate and barium hydroxide. The principal use is in pyrotechnics, but there are also small uses in the curing of polysulfide rubbers and in the production of certain titanium–aluminum alloys.

Barium peroxide is a strong oxidizer and can cause fire when in contact with combustible materials. It is a powerful irritant to skin, eyes, and mucous membranes (2). Consequently, it is also toxic via the subcutaneous route; protective clothing should be worn during handling. The LD₅₀ value (mouse, oral) is 50 mg/kg (2).

This product has been assigned UN No. 1449 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances. It is made by Solvay Deutschland, which uses the trade name IXP_{ER} 90 B, and by L'Air Liquide. The U.S. price in 1994 was \$12/kg; the French price was Fr. 75/kg.

Group 12 (IIB) Peroxides

Zinc Peroxide. Zinc peroxide [1314-22-3], ZnO₂, is a yellow solid, generally similar to magnesium peroxide. The commercial product is a pale yellow powder containing about 55% ZnO₂ and 9% active oxygen. It is stable at room temperature, decomposes at above 150°C to oxygen and zinc oxide, and decomposes explosively at 190–212°C. It is stable in dry air but loses its oxygen in moist air and on heating. It is insoluble in water but dissolves in dilute acid, liberating hydrogen peroxide.

Zinc peroxide is produced commercially by drying a slurry of zinc oxide in hydrogen peroxide. It also can be produced by the reaction of alkaline hydrogen peroxide and a soluble zinc salt (10). It is used as an accelerator in rubber-compounding, as a curing agent for synthetic elastomers, and as a deodorant for wounds and skin diseases. Zinc peroxide is a powerful irritant to skin, eyes, and mucous membranes (2); protective clothing should be worn during handling. The systemic toxicity is similar to that of zinc oxide (2), for which the LD₅₀ (rat, oral) is 7950 mg/kg (2).

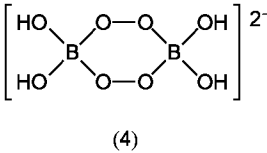
Zinc peroxide is a strong oxidizer and can cause fire when in contact with combustible materials. It has been assigned UN No. 1516 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances. It is made by Solvay Deutschland, which uses the trade name IXP_{ER} 55 Z, and by L'Air Liquide. In 1994, the price in France was Fr. 65–80/kg (\$7–9/kg).

Group 13 (IIIB) Peroxides

Boron Compounds.

Nomenclature. The naming of sodium perborate, one of the most important commercial boron compounds, has long been confused. The stoichiometry of the most important hydrate is NaBO₃·4H₂O, often called sodium perborate tetrahydrate. Only one of the oxygen atoms acts as an oxidant, however, and the formula has often been written as NaBO₂·H₂O₂·3H₂O, and the compound referred to as a trihydrate. The crystal structure, established in 1961, contained the cyclic diperoxodiborate anion [B₂(O₂)₂(OH)₄]²⁻ and six associated molecules of water. Thus the compound has more recently come to be called the hexahydrate. The crystallographically derived names are used to avoid confusion (11,12). The commercial or common names are also given. The prefix *per* usually signifies an element in its highest valency state. Because the boron is always trivalent, the prefix *peroxo*, recommended by IUPAC, is used herein.

Sodium Peroxoborate Hexahydrate. The compound sodium peroxoborate hexahydrate [10486-00-7] (sodium perborate tetrahydrate), Na₂[B₂(O₂)₂(OH)₄]·6H₂O, was formerly written as NaBO₃·4H₂O. This material has been an important commercial bleaching agent for many years (see BLEACHING AGENTS). It is produced by mixing hydrogen peroxide with an alkaline solution of sodium borate made from a boron mineral, usually native borax (Rasorite or Tincal) (see BORON COMPOUNDS). The mineral is dissolved in aqueous sodium hydroxide and the solution is clarified; hydrogen peroxide and a stabilizer such as magnesium silicate are then added. The product crystallizes when cooled to ambient temperature and is removed by filtration or centrifuge. It usually contains several percent of moisture and is dried by warm air at a temperature below 60°C. There are a number of minor variations on this basic process. It is possible to control the crystal morphology or the state of aggregation of the product in order to improve handling characteristics. Because x-ray diffraction (11,12) shows the crystal to contain a dimeric cyclic anion (4), the systematic name is disodium tetrahydroxo-di-μ-peroxo-diborate(III) hexahydrate [10486-00-7].



The commercial product is a white, crystalline powder having an active oxygen content of at least 10%. It melts at about 60°C; however, if water vapor is free to escape during heating, the crystals do not melt but are converted to the anhydrous peroxoborate.

In aqueous solution, all the sodium peroxoborates dissociate for the most part into boric acid, or its anion, and hydrogen peroxide. Peroxoborate species are also present in these solutions, depending on the pH and the concentration for the species type. The nature of these species has been extensively examined by classical physicochemical methods (13), by nmr, and by Raman spectroscopy (14–17). Both monomeric and polymeric species are usually present. There is some evidence (18) suggesting that these peroxoborates are more reactive than hydrogen peroxide alone under similar conditions.

Sodium peroxoborate hexahydrate is an important ingredient of many household detergents, working best at temperatures above 60°C. It is also used in dishwasher detergents, denture cleaners, as well as foot and bath salts. The textile industry generally uses hydrogen peroxide for bleaching, but there are a few areas in which sodium peroxoborate hexahydrate is preferred.

Organic chemists have been using sodium peroxoborates as oxidants since the 1980s (19). The hexahydrate and monohydrate generally behave similarly. In the first examples of such oxidations, published in 1983 (20), a solution of the hexahydrate in glacial acetic acid was found to oxidize aromatic amines to nitroarenes, and sulfides to sulfoxides or sulfones (21). Since that time, many other such oxidation systems have been studied, although the identity of the oxidants in such systems remains unclear. The oxidant was originally considered to be a peroxyacetic acid, but kinetic studies have been inconsistent with this hypothesis. At the conclusion of these reactions in acetic acid, a white solid commonly precipitates; this precipitate is not a peroxygen compound but appears to be a sodium tetraborate having associated acetic acid (19). Other organic solvents, or phase-transfer systems, can be used in place of acetic acid. There have been several mechanistic studies of organic oxidations by peroxoborate in water (22–24).

The toxicity of sodium peroxoborate hexahydrate in solution is equivalent to those of sodium borate and hydrogen peroxide. The LD₅₀ (mouse, oral) is 1060 mg/kg (2). Local use of high concentrations in the mouth can cause chemical burns and other problems (25). No TLV has been established.

The product is considered nonhazardous for international transport purposes. However, it is an oxidizing agent sensitive to decomposition by water, direct sources of heat, catalysts, etc. Decomposition is accompanied by the liberation of oxygen and heat which can support combustion and cause pressure bursts in confined spaces. Decomposition in the presence of organic material is rapid and highly exothermic.

Sodium peroxoborate hexahydrate is made commercially by Atochem (France), Belinka (Slovenia), Ausimont (Italy), Caffaro (Italy), Degussa (Germany, United States), Eka Nobel (Sweden), Etibank (Turkey), Foret (Spain), L'Air Liquide, Solvay Interlox (Australia, Belgium, France, Germany, Italy, Portugal, Spain, United Kingdom, United States), and Treibacher Chemische Werke (Austria). In 1994, the total world manufacturing capacity for the

hexahydrate was about 900,000 t, about one-third of which was for conversion to sodium peroxoborate (sodium perborate monohydrate). The U.S. price in 1994 was \$0.79–0.86 kg, the German price was DM 0.75–1.00 kg (\$0.63–0.57 kg).

Sodium Peroxoborate Tetrahydrate. The compound sodium peroxoborate tetrahydrate [28962-65-4] (sodium perborate trihydrate), Na₂B₂(O₂)₂[(OH)₄] ·4H₂O, was formerly written as NaBO₃ ·3H₂O. A number of procedures have been published (13) for producing this hydrate; most involve the crystallization from water at 40–50°C. The crystal structure (26) shows the presence of the same cyclic perborate anion (4) as that in the hexahydrate.

Sodium peroxoborate tetrahydrate is the most stable of the three peroxoborate hydrates under ambient conditions. It has, however, never been commercialized because it is slow when dissolving in water.

Sodium Peroxoborate. Sodium peroxoborate [10332-33-9] (sodium perborate monohydrate), Na₂[B₂(O₂)₂(OH)₄], formerly written as NaBO₃ ·H₂O, is known only as a microcrystalline powder, made by dehydrating the hexahydrate. The crystal structure has not been determined, but the vibrational spectrum (27) indicates the presence of the same cyclic peroxodiborate anion (4) as that in the hexahydrate as well as in the tetrahydrate.

The commercial product has an active oxygen content of at least 15%. This product has replaced the hexahydrate in some household detergents and other domestic products because it dissolves faster and has a greater content of active oxygen per unit volume of granular product.

The toxicity of sodium peroxoborate is similar to that of the hexahydrate. The LD₅₀ (cat, intravenous) is 600 mg/kg; the LD₅₀ (rabbit, intravenous) is 78 mg/kg (28). Sodium peroxoborate is a severe eye irritant, but not a skin irritant. Absorption through large areas of abraded or damaged skin can give systemic boron poisoning (2). The maximum eight-hour time-weighted average exposure is 5 mg/m³ (2).

The product is considered nonhazardous for international transport purposes. However, it is an oxidizing agent sensitive to decomposition by water, direct sources of heat, catalysts, etc. Decomposition of sodium peroxoborate is accompanied by the liberation of oxygen and heat which can support combustion and cause pressure bursts in confined spaces. Decomposition in the presence of organic material is rapid and highly exothermic.

The world market in 1993 was about 450,000 t. Sodium peroxoborate is made commercially by Atochem, Ausimont, Caffaro, Degussa, Eka Nobel, Foret, L'Air Liquide, Solvay Interox (Belgium, Italy, Germany, United Kingdom, United States), and Treibacher Chemische Werke. The U.S. price in 1994 was in the range \$1.32–1.43/kg, the German price was DM 1.35–1.60/kg (\$0.78–0.92/kg).

Anhydrous Sodium Perborate. Anhydrous sodium perborate [7632-04-4], NaBO₃, is an ill-defined, powdery material, made by heating sodium peroxoborate in a current of dry air in a fluidized bed at 150–160°C. Attempts to characterize this material physically have been largely unsuccessful (29,30). Its electron spin resonance spectrum shows several types of free radicals in the powder. It should perhaps be regarded more as an amorphous assemblage of radicals than as a defined compound.

Anhydrous sodium perborate effervesces in water. The commercial product contains a minimum of 13% oxygen which can be released in this way; however, the content of active oxygen that is released into solution in the water is only about 2%. It is used mainly as an ingredient in denture-cleaning formulations. No toxicological data have been reported on this product, except that in humans, swallowing large amounts can cause nausea, vomiting, and diarrhea. Anhydrous sodium perborate is irritating to eyes, skin, and mucous membranes. It is also mutagenic to *E. coli* (2).

This product has been assigned UN No. 3247 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances. It is made commercially by L'Air Liquide and Degussa (Germany), the latter sells it under the trade name Oxoborate. The U.K. price in 1994 was £1.90/kg (\$2.81/kg).

Group 14 (IVB) Peroxides

Peroxocarbonates. Peroxocarbonates contain the C–O–O– group and should be distinguished from the carbonate peroxohydrates. Although no crystal structures have been determined, the nature of the peroxocarbonates has been deduced from vibrational spectra (31). These compounds can be prepared by three general methods: reaction of carbon dioxide and a solution of the metal hydroxide in hydrogen peroxide, anodic oxidation of normal carbonates at low temperatures, and oxidation of aqueous solutions of carbonates with elemental fluorine. Only the peroxocarbonates of the alkali metals are known. Some are peroxomonocarbonates, containing the CO^{2–4} anion; others are peroxodicarbonates, containing the C₂O^{2–} anion. Mixed alkali metal hydrogen salts are also known. Those salts that have been reasonably well-characterized include Li₂(CO₄) ·H₂O, Na₂(C₂O₂) ·nH₂O, K₂(C₂O₂), Rb₂(C₂O₂), Cs₂(C₂O₂), NaHCO₄ ·H₂O, KHCO₄, and RbHCO₄.

There are international transport regulations controlling the transport of sodium percarbonate, which assigned it to Class 5.1, oxidizing substances, however, no such compound has ever been commercialized, and sodium carbonate peroxohydrate is treated as nonhazardous. The origin of this item is not known.

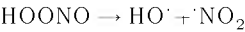
Peroxosilicates. No solid peroxosilicates are known. There is some indirect evidence based on catalytic studies for the existence of peroxosilicates in solution (32,33), but nmr spectra of aqueous solutions of silicates mixed with hydrogen peroxide give no evidence for peroxo species (34). A peroxohydrate of sodium silicate, Na₃SiO₃ ·3H₂O₂, is well established (35–37) but has found no application.

Peroxotin Compounds. Older literature (38) records some tin peroxides or peroxohydrates, but these claims have not been substantiated. In contrast, organometallic peroxotin compounds are well established (39).

Group 15 (VB) Peroxides

Peroxonitrous Acid and Its Salts. Peroxonitrous acid [14691-52-2], HOONO14691-52-2, is an isomer of nitric acid, HNO₃, to which it rapidly converts. The half-life of peroxonitrous acid at 0°C is 10 s; at 27°C, 0.23 s (40,41). It has been known since 1904 (42) that the yellow solution made by mixing nitrous acid and hydrogen peroxide at low temperature contains a stronger oxidant than either ingredient alone, but the chemistry involved was not put on a sound basis until 1994 (43). Additional preparatory methods are also available.

Peroxonitrous acid can decompose by two pathways: isomerization to nitric acid, and dissociation into the hydroxyl radical and nitrogen dioxide.



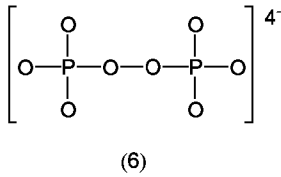
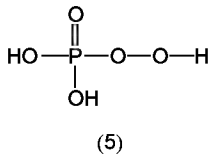
The hydroxyl radical is responsible for some of the oxidation products of organic compounds by peroxonitrous acid.

The peroxonitrite ion, OONO[–], an isomer of the nitrate ion, NO₃[–], is fairly stable under certain conditions. It can be made by rapidly neutralizing cold peroxonitrous acid. It can also be made by irradiating sodium nitrate crystals with uv, x-rays, γ-rays, neutrons, or electrons. The irradiated crystals are yellow, and in a closed bottle can retain the peroxonitrite for months. When the crystals are dissolved in water at pH 12 that contains a chelating agent to sequester metal impurities, a yellow, fairly stable solution of peroxonitrite results. Acidification of this solution generates the free peroxyacid, which then decomposes with the liberation of hydroxyl radicals. This system has been suggested as a convenient source of aqueous hydroxyl radicals.

Peroxonitrite is believed to be present in the crystals of nitric acid trihydrate that form in the stratosphere and in Martian soil (see EXTRATERRESTRIAL MATERIALS). Peroxonitrous acid may be present in mammalian blood and other biochemical systems. However, peroxonitric acid, HNO₄, is not known.

Before the chemistry of peroxonitrous acid was understood, these two acids were sometimes confused.

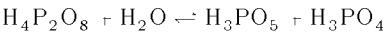
Peroxophosphoric Acids and Their Salts. In its usual impure form (H₃PO₄ is the main contaminant), peroxomonophosphoric acid [13598-52-2] (5), is a viscous, colorless liquid. The three ionization constants for peroxomonophosphoric acid are pK₁ = 1.1, pK₂ = 5.5, and pK₃ (peroxide proton) = 12.8 (44). Oxidations comparable to those of peroxomonosulfuric acid, H₂SO₅, occur in acid solutions of ca pH 2, but at higher pH values, H₃PO₅ becomes less reactive as an oxidant and more unstable with respect to decomposition (44). The structure of H₃PO₅ is probably similar to that of H₂SO₅.



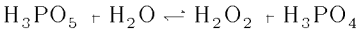
Peroxomonophosphoric acid can be prepared by the hydrolysis of peroxodiphosphates in aqueous acid and by the reaction of hydrogen peroxide with phosphorus pentoxide (45). It is not produced or used commercially and the salts that have been prepared are unstable and impure.

Pure peroxodiphosphoric acid [13825-81-5], H₄P₂O₈, has not been obtained, but its properties in aqueous solution are understood. As indicated by ir and Raman spectral data, the P₂O⁴₈ anion (6) consists of two anionic PO₄ tetrahedrons joined by an –O–O– bond (46).

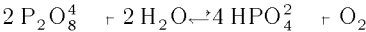
There are four ionizable hydrogen atoms on the corresponding peroxodiphosphoric acid, for which p*K*₁ = 0.3, p*K*₂ = 0.5, p*K*₃ = 5.2, and p*K*₄ = 7.7 (47). The rates of hydrolysis of the peroxodiphosphate ion are dependent on both pH and temperature, eg, as the hydrogen ion concentration increases, the rate increases (48). Peroxodiphosphoric acid hydrolyzes to peroxomonophosphoric acid according to the following equation:



This reaction takes place ca 50 times more rapidly than the one in which peroxomonophosphoric acid hydrolyzes to hydrogen peroxide:



The rates of hydrolysis for the peroxophosphoric acids are more rapid than the corresponding reactions of the peroxosulfuric acids. The peroxodiphosphate ion is extremely resistant to decomposition by oxidation of water:



However, the peroxomonophosphate ion decomposes relatively rapidly in aqueous solution. A mixture of peroxodiphosphoric and peroxomonophosphoric acids can be produced by treating a cold phosphoric acid solution with elemental fluorine (qv) (49). Peroxodiphosphoric acid is not produced commercially. Ammonium, lithium, sodium, potassium, rubidium, cesium, barium, zinc, lead, and silver salts have all been reported. The crystal structures of the ammonium, lithium, sodium, and potassium compounds, which crystallize with varying numbers of water molecules, have been determined (50).

Tetrapotassium peroxodiphosphate [15593-49-9], K₄P₂O₈, is a colorless, crystalline solid, soluble in water to 42.2 wt % at 0°C and 51.2 wt % at 40°C (51). The pH of a 2 wt % aqueous solution is ca 9.6. It does not melt on heating but decomposes exothermally at 387°C; Δ*H* = 29.6 kJ mol⁻¹ (–123.8 kcal mol⁻¹). It loses oxygen and forms potassium pyrophosphate. Tetrapotassium peroxodiphosphate is stable in alkaline solutions but hydrolyzes to the peroxomonophosphate in strongly acidic solutions. It is a strong oxidizing agent with an oxidation potential of 2.07 V. However, it is kinetically inhibited and does not readily oxidize many organic materials.

Tetrapotassium peroxodiphosphate is produced by electrolysis of a solution containing dipotassium phosphate and potassium fluoride (52). Alkalinity favors the formation of the P₂O⁴₈ anion, whereas the PO³₅ anion is produced in larger yields in acidic solution. It is therefore possible to obtain an 80% yield of K₄P₂O₈ by choosing the proper conditions. The tetrapotassium peroxodiphosphate can be crystallized from solution by evaporation of water to form a slurry. The crystals can be separated from the slurry and dried. The material is noncorrosive and cannot be catalytically decomposed by iron ions.

In the early 1970s, FMC (53,54) developed textile desizing processes based on tetrapotassium peroxodiphosphate, which at that time was available in development quantities. This peroxodiphosphate was chosen in place of the more usual peroxodisulfate because the former is active at a higher temperature. However, the desizing processes were not adopted and the salt has never been commercialized. Tetrapotassium peroxodiphosphate is being investigated as an ingredient in toothpaste as an anticalculus agent and bactericide (55). The similarity of the chemistry of the peroxodiphosphates to that of the peroxodisulfates, and the formers' slower rates of oxidation, have prevented the peroxodiphosphates from becoming useful commercial products.

Arsenic Peroxides. Arsenic peroxides have not been isolated; however, elemental arsenic, and a great variety of arsenic compounds, have been found to be effective catalysts in the epoxidation of olefins by aqueous hydrogen peroxide. Transient peroxoarsenic compounds are believed to be involved in these systems. Compounds that act as effective epoxidation catalysts include arsenic trioxide, arsenic pentoxide, arsenious acid, arsenic acid, arsenic trichloride, arsenic oxychloride, triphenyl arsine, phenylarsonic acid, and the arsenates of sodium, ammonium, and bismuth (56). To avoid having to dispose of the toxic residues of these reactions, the arsenic can be immobilized on a polystyrene resin (57).

Group 16 (VIB) Peroxides

Peroxosulfuric Acids and Their Salts. Two kinds of peroxosulfuric acid are known: peroxomonosulfuric and peroxodisulfuric acids. Neither is available commercially in the pure state. The name Caro's acid is commonly used as a synonym for peroxomonosulfuric acid, but is better reserved for the equilibrium mixture with sulfuric acid.

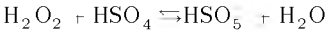
Peroxomonosulfuric acid [7722-86-3], H₂SO₅, when pure, forms colorless crystals that melt with decomposition at 45°C. One of its protons is strong, as in sulfuric acid, but its other proton, which is on the peroxide group, is weak (p*K*_a = 9.4). Peroxomonosulfuric acid is a strong oxidizing agent:

$$\text{H}_2\text{SO}_5 + 2 \text{H}^+ + 2 e^- \xrightarrow{E^0 = 1.81 \text{ V}} \text{H}_2\text{SO}_4 + \text{H}_2\text{O}$$

The correct value for the oxidation potential of the HSO₅–HSO₄ couple was established by a thermodynamic method in 1979 (58) and confirmed in 1982 (59). Previously published values have been found inaccurate.

Peroxomonosulfuric acid oxidizes cyanide to cyanate, chloride to chlorine, and sulfide to sulfate (60). It readily oxidizes carboxylic acids, alcohols, alkenes, ketones, aromatic aldehydes, phenols, and hydroquinone (61). Peroxomonosulfuric acid hydrolyzes rapidly at pH <2 to hydrogen peroxide and sulfuric acid. It is usually made and used in the form of Caro's acid.

Caro's Acid. Caro's acid is named after Heinrich Caro (1834–1910), who first described its preparation and oxidizing properties in 1898. Herein Caro's acid is used to designate the equilibrium mixtures that result from mixing hydrogen peroxide and sulfuric acid. These liquids mix instantly, generating a considerable amount of heat. The equilibrium constant for this reaction is 0.1 (62).



Because the product is decomposed by heat, it is essential either to remove the heat of reaction quickly or to use the product quickly. The first option is known as the isothermal process; the second option, perfected and commercialized in the early 1990s (63,64), is known as the adiabatic process.

The materials of construction for the mixing device and storage vessels must be selected carefully. Glass (qv), polytetrafluoroethylene, or certain kinds of stainless steels are usually used. Glass must be pickled with nitric acid before use.

In a typical isothermal process, 70° hydrogen peroxide is added to 98° sulfuric acid, and subjected to rapid stirring and efficient cooling, so that the temperature does not rise to above 15°C. If equimolar quantities of reactants are used, the product contains 42° H₂SO₅ and 10° H₂O₂. Although the reaction may seem simple, many of its features are critically important and it should only be attempted following advice from specialists.

In the adiabatic process, the reactants are mixed rapidly in a small-volume, high throughput static mixer without cooling (65). The hot product is used directly. A 250-mL reactor of this type can produce 8 t d of peroxomonosulfuric acid.

Caro's acid is finding increasing application in hydrometallurgy, pulp bleaching, effluent treatment, and electronics. There are several applications of Caro's acid in hydrometallurgy. It is usually made on-site by either the isothermal or the adiabatic process. The latter method is preferred because its capital cost is less and the system is safer due to the fact that the product is used as soon as it is made.

Caro's acid has been used in Australia as an oxidant in the acid-leaching of uranium ores. It acts by oxidizing the iron present in the solution from Fe²⁺ to Fe³⁺. This Fe³⁺ then oxidizes the uranium. Alternative oxidants that have been used include pyrolusite and chlorate ion. These are both undesirable because their effluents, containing Mn³⁺ or Cl⁻, contaminate watercourses.

Another hydrometallurgical use of Caro's acid is for destroying cyanide effluents from gold-leaching. It is more useful than catalyzed hydrogen peroxide for this duty because it also destroys thiocyanate ion. This application has been demonstrated in the laboratory (66) and at two gold mines (66,67).

A proprietary form of Caro's acid is sold to the electronics industry under the trade name Nanostrip. Used for reclaiming defective silicon wafers, it is manufactured by Cyantech (United States), Micro-Image Technology (UK), and RASA Industries (Japan).

Caro's acid is effective in delignifying wood pulp (qv) made by chlorine-free bleaching sequences. When conditions are carefully controlled, the mechanical properties of the final paper (qv) are not impaired. These processes were developed in the 1980s and commercialized in the 1990s (68).

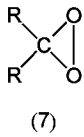
Caro's acid is highly corrosive and a powerful oxidant. Its acidic properties are similar to those of sulfuric acid of equivalent strength. A strong irritant, it is toxic and should always be handled accordingly. No specific toxicological data are available.

Depending on the strength of the product, Caro's acid should be transported in accordance with the relevant regulations pertaining to the most appropriate sulfuric acid solution or to those of liquid oxidizers not otherwise specified (NOS).

Peroxomonosulfates. When oleum is mixed with hydrogen peroxide and the mixture is partially neutralized by potassium hydroxide, a triple salt [37222-66-5] crystallizes out. In the old nomenclature, the formula for the triple salt was written as if it comprised three salts: 2 KHSO₅·KHSO₄·K₂SO₄, hence the name. It has also been described as a mixture of the three salts, which it is not. The formula should be written in the more general form as K₅(HSO₅)₂(HSO₄)(SO₄). The crystal structure is disordered and not fully established, but reports on it have appeared (69,70). The presence of the hydrogen monoperoxosulfate ion HSO₅⁻ in the crystal is also mentioned.

The commercial product is a white, finely crystalline powder containing a minimum of 4.7° active oxygen. It is used because it is stable and safe, despite being a powerful oxidant. Its main use is in denture cleaners that function without mechanical assistance, in which discolorations are bleached and organic deposits are oxidized. It is also used in dishwashing detergents and toilet bowl cleaners. Furthermore, it is used in the metal-fabricating industry as a mild etchant and pickling agent (see METAL SURFACE TREATMENTS), and in the electroplating (qv) industry for detoxifying cyanide solutions. In the latter application, it is mixed with hydrogen peroxide. Finally, it is used in the textile industry for rendering wool (qv) shrink-resistant and nonfelting.

In general, peroxomonosulfates have fewer uses in organic chemistry than peroxodisulfates. However, the triple salt is used for oxidizing ketones (qv) to dioxiranes (7) (71,72), which in turn are useful oxidants in organic chemistry. Acetone in water is oxidized by triple salt to dimethyldioxirane, which in turn oxidizes alkenes to epoxides, polycyclic aromatic hydrocarbons to oxides and diones, amines to nitro compounds, sulfides to sulfoxides, phosphines to phosphine oxides, and alkanes to alcohols or carbonyl compounds.



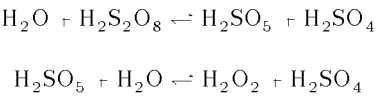
No toxicological studies have been reported on the triple salt. However, because of the common confusion of this compound with potassium hydrogen monoperoxosulfate monohydrate, it is possible that the published descriptions of the toxic properties of this latter compound actually refer to the triple salt. If this is so, then the triple salt must be regarded as toxic and irritating to skin, eyes, and mucous membranes (2).

The triple salt is classified by the UN not as an oxidizer but as a corrosive, and thus must be transported under the UN No. 1759 for corrosive solids NOS. It should be kept away from combustible material.

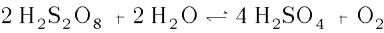
The triple salt is better known by the trademarks Caroat (Degussa), OXONE Monopersulfate Compound (Du Pont), and Curox (Laporte). It is also known as potassium caroate. It has been made on a commercial scale since the 1950s, and the world market in 1994 was several thousand tons. It is made commercially by Peroxid-Chemie (Germany), Degussa (Germany), Du Pont (United States), and Migas (Japan). In 1994, the United Kingdom price was J1.80/kg (\$2.67/kg).

Potassium hydrogen monoperoxosulfate monohydrate [14696-73-2], KHSO₅·H₂O, related to the triple salt, is not made commercially. The crystal structure has been determined and some features of its Raman and ir spectra recorded (69). This compound is more stable under x-rays than the triple salt. The O—O distance is 0.1460 nm. The dihedral angle of the O—O moiety is about 90°, similar to that in solid hydrogen peroxide. This compound is reported as toxic and irritating to eyes, skin, and mucous membranes (2). Although undoubtedly correct, this description probably better relates to the triple salt.

Peroxodisulfuric Acid. Also called persulfuric acid, and Marshall's acid (73), peroxodisulfuric acid [13445-49-3], H₄S₂O₈, when pure, forms colorless crystals that melt with decomposition at 65°C. The structure of the S₂O₈²⁻ ion has been established by x-ray analyses of the cesium and ammonium salts as two anionic SO₄ tetrahedra joined by an O—O linkage (74). Peroxodisulfuric acid is a strong acid but not stable. It is seldom isolated but is synthesized and used in solution. Solutions of peroxodisulfuric acid are reasonably stable when cool but hydrolyze rapidly to hydrogen peroxide and sulfuric acid when heated in strongly acidic solutions. This hydrolysis proceeds stepwise and involves the intermediate peroxomonosulfuric acid:



A slower decomposition reaction also occurs:



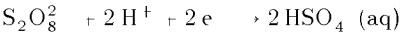
Peroxodisulfates. The salts of peroxodisulfuric acid are commonly called persulfates, three of which are made on a commercial scale: ammonium peroxodisulfate [7727-54-0], (NH₄)₂S₂O₈; potassium peroxodisulfate [7727-21-1], K₂S₂O₈; and sodium peroxodisulfate [7775-27-1], Na₂S₂O₈. The peroxodisulfates are all colorless, crystalline solids, stable under dry conditions at ambient temperature but unstable above 60°C.

All the peroxodisulfates are made commercially by electrolytic processes (75). The classical method was developed from the obsolete Weissenstein process (4) for making hydrogen peroxide, in which a solution containing ammonium sulfate and sulfuric acid was electrolyzed between platinum electrodes. Ammonium peroxodisulfate was crystallized directly from the spent electrolyte, and the other salts were made from this by adding alkalis and volatilizing the ammonia. In the 1970s, an improved electrolytic process was developed (76), giving all three peroxodisulfate salts directly from the

corresponding hydrogen sulfate salts; one multipurpose cell could thus produce whichever salt was required. Cells of the old design used diaphragms of asbestos that were troublesome to maintain and would no longer be approved for safety reasons. Cells of the new design operate continuously and do not use a diaphragm.

In addition to cost, several factors influence the choice of salt for a particular application. The ammonium salt is the most soluble in water, but for some applications the presence of the ammonium ion may be undesirable. The sodium salt is almost as soluble as the ammonium salt at ambient temperatures and above. The potassium salt is much less soluble.

The peroxodisulfate ion in aqueous solution is one of the strongest oxidizing agents known. The standard oxidation–reduction potential for the following reaction is 2.08 V (77,78).



Reactions involving the peroxodisulfate ion are usually slow at ca 20°C. The peroxodisulfate ion decomposes into free radicals, which are initiators for numerous chain reactions. These radicals act either thermally or by electron transfer with transition-metal ions or reducing agents (79).



The principal use of the peroxodisulfate salts is as initiators (qv) for olefin polymerization in aqueous systems, particularly for the manufacture of polyacrylonitrile and its copolymers (see ACRYLONITRILE POLYMERS). These salts are used in the emulsion polymerization of vinyl chloride, styrene–butadiene, vinyl acetate, neoprene, and acrylic esters (see ACRYLIC ESTER POLYMERS; STYRENE; VINYL POLYMERS).

Etching of printed circuit boards and removal of photoresists are also important applications (see ELECTRONIC MATERIALS; INTEGRATED CIRCUITS). Bleaching of textiles and natural fibers and finishing of furs are both long-established applications. Other established applications include curing grouts for soil stabilization (qv), initiating polymerization of graphite filament coatings, cleaning metal surfaces prior to plating or adhesive bonding, and regenerating active carbon.

An expanding development is the use of peroxodisulfates as oxidants in organic chemistry (80,81). These reactions are initiated by heat, light, gamma rays, or transition-metal ions. The primary oxidizing species is usually the sulfate ion radical, $\text{SO}_4^{\cdot -}$, but the peroxodisulfate anion ($\text{S}_2\text{O}_8^{2-}$) and peroxomonosulfuric acid may also be involved. The metal ions commonly used in association with peroxodisulfates are Ag^+ , Cu^+ , Fe^{2+} , Ce^{3+} , Mn^{3+} , and Ti^{3+} . Most organic compounds can be oxidized by these systems, and a great variety of products can be obtained, some in good yield. Phenols yield *p*-hydroquinones, aromatic amines yield *o*-aminophenols, and alkyl aromatic hydrocarbons yield aromatic aldehydes. Many other examples are described in the literature (80,81).

The three peroxodisulfates are all toxic and irritating to skin, eyes, and mucous membranes. Published toxicity studies are as follow (28):

Peroxodisulfate	Animal system	LD ₅₀ , mg/kg
ammonium	rat, oral	689
potassium	rat, oral	802
sodium	mouse, oral	226

The LD_{LO} value for sodium peroxodisulfate using iv administration in rabbits is 178 mg/kg.

The U.K. maximum occupational exposure level in air for each of the three salts is 1 mg/kg; these salts should be handled only when wearing suitable protective clothing. The peroxodisulfates assist combustion by releasing oxygen, and must be stored away from combustible materials. Contamination by rust or traces of many metals can cause catalytic decomposition. The peroxodisulfates should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances.

The peroxodisulfates are made commercially by Akkim (Turkey), Degussa (Germany), FMC, Migas, Peroxid-Chemie, ERB Engineering (India), Sangen (Taiwan), and Tokai Denka (Japan). In 1994 the U.S. price of the sodium salt was \$1.92/kg, the potassium salt \$2.07/kg, and the ammonium salt \$1.68/kg.

Other Metal Peroxides

Transition-Metal Peroxides. A 1964 review paid tribute to the significance of transition-metal peroxide and peroxo chemistry for the catalysis of oxidations and the storage and use of oxygen in biological systems (82). Since that time, many more inorganic peroxo compounds have been isolated and several more transition-metal-catalyzed organic reactions have been commercialized (83–88). However, transition-metal peroxides, as isolated species, have no place in chemical technology because they are too dangerously explosive.

Transition metals can be divided into two groups according to the characteristics of their peroxides. The first group comprises those metals that, in their highest oxidation states, have no *d* electrons, eg, Ti^{4+} and W^{6+} . These metals form peroxides from hydrogen peroxide, the colors of which result from charge-transfer between the peroxide group and the metal ion. The peroxo species act as electrophiles.

The other group of transition metals comprises those metals that retain *d* electrons in their normal valence states, eg, Co^{3+} and Pt^{2+} . These metals form peroxides from dioxygen or from hydrogen peroxide. Their colors result from *d*–*d* transitions. These peroxo species act as nucleophiles.

Transition-metal-catalyzed oxidations may or may not proceed via peroxocomplexes. Twelve important industrial organic oxidation processes catalyzed by transition metals, many of which probably involve peroxo intermediates, have been tabulated (88). Even when peroxo intermediates can be isolated from such systems, it does not necessarily follow that these are true intermediates in the main reaction.

Actinide Peroxides. Many peroxo compounds of thorium, protactinium, uranium, neptunium, plutonium, and americium are known (82,89). The crystal structures of a number of these have been determined. Perhaps the best known are uranium peroxide dihydrate [*1344-60-1*], $\text{UO}_4 \cdot 2\text{H}_2\text{O}$, and, the uranium peroxide tetrahydrate [*15737-4-5*], $\text{UO}_4 \cdot 4\text{H}_2\text{O}$, which are formed when hydrogen peroxide is added to an acid solution of a uranyl salt.

Uranium peroxide has found several applications in the nuclear energy industry. It provides a method for precipitating uranium from solution without introducing any extraneous cations or anions. It has been used in the extraction of uranium from its ores, where ammonia and hydrogen peroxide are used to precipitate uranium from leachate. It is also used in some fuel cycles (90). When the homogeneous aqueous nuclear reactor was being developed (91), there was a concern that uranium peroxide, formed from the autogenous hydrogen peroxide, might precipitate from solution. Under certain conditions this might have happened, but the project was abandoned for other reasons (see NUCLEAR REACTORS).

Peroxohydrates

Peroxohydrates are crystalline adducts containing molecular hydrogen peroxide. These are commonly called perhydrates, but this name is better avoided because *per* historically implied the maximum oxidation state and *hydrate* implies the presence of water, neither of which apply to peroxohydrates. They have also been called hydroperoxidates (92).

On dissolution in water, peroxohydrates liberate hydrogen peroxide into solution. Some peroxo salts also liberate hydrogen peroxide when dissolved in water, and before the introduction of x-ray crystallography, compounds within these classes were often confused with each other.

Peroxohydrates are usually made by simple crystallization from solutions of salts or other compounds in aqueous hydrogen peroxide. They are fairly stable under ambient conditions, but traces of transition metals catalyze the liberation of oxygen from the hydrogen peroxide. Early work on peroxohydrates has been reviewed (92).

Sodium Carbonate Peroxohydrate. Known commercially as sodium percarbonate, sodium carbonate peroxohydrate [15630-89-4] does not contain the C—O—O—C group and is not a peroxocarbonate. The stoichiometry is $2\text{Na}_2\text{CO}_3 \cdot 3\text{H}_2\text{O}_2$. The material is made commercially by three processes: batch crystallization, continuous crystallization, and fluid-bed reaction (93). The crystallization processes are similar to the processes for making sodium peroxoborate hexahydrate. Solutions of hydrogen peroxide and sodium carbonate are mixed in proportions close to the stoichiometry of the end product. Stabilizers are added and the product is separated by cooling and the use of salting-out agents such as sodium chloride. In the fluid-bed process, hydrogen peroxide solution and sodium carbonate solution are injected into a fluid bed of the product fed by warm air. The bed serves as a combined reactor, granulator, and drier. Stabilizers are usually introduced with the hydrogen peroxide.

The commercial product is a white powder containing a minimum of 13% of active oxygen and up to 15% of anhydrous sodium carbonate. The solubility in water at 20°C is about 150 g/L.

The crystal structure (94) shows the hydrogen peroxide to be rather loosely bound, so that the H_2O_2 exerts a small vapor pressure above the solid. It is easily displaced by moisture with disruption of the crystal structure. Because the compound is alkaline (pH 10–11 in solution) and hydrogen peroxide is unstable in alkali, the hydrogen peroxide liberated from the peroxohydrate tends to decompose as it escapes. The peroxohydrate is storage-stable if dry, but unstable if wet. The ready availability of hydrogen peroxide from the peroxohydrate, often an advantage in synthetic applications, can lead to violent exothermic reactions if the peroxohydrate is mixed with an excess of an oxidizable substance.

When dissolved in water, the solution is identical with that obtained by dissolving sodium carbonate in aqueous hydrogen peroxide. There is some evidence for the presence of the traces of true peroxocarbonate anion, HCO_4^- , in these solutions (95). If the peroxohydrate is heated for about an hour at 100°C and then allowed to cool to room temperature, some decomposition occurs and the product effervesces when placed in water. Electron spin resonance experiments (64) indicate that free radicals are present in this partially decomposed material, but the nature of these radicals is obscure.

The principal use of sodium carbonate peroxohydrate is as a bleaching agent in domestic and laundry detergents. It dissolves rapidly in water and offers the advantage of a high concentration of active oxygen per unit volume of granulated product. It is used also for industrial textile-bleaching, tripe-bleaching, and in denture cleaners. It can also be used as a convenient oxidant in organic chemistry.

In general, the compound is not as stable in domestic detergent formulations as sodium peroxoborate hydrates. Active oxygen tends to be lost over a period of months, the rate of loss depending on the other ingredients present. A common detergent ingredient detrimental to the stability of the peroxohydrate is zeolite A, which is used to sequester calcium and magnesium ions in hard water. It is believed that hydrogen peroxide vapor from the peroxohydrate enters the zeolite pores and is catalytically decomposed there. A number of patents describe processes for stabilizing the peroxohydrate against such zeolite-induced decomposition. One of these (96) recommends coating the crystals of the peroxohydrate with various inorganic salts, including borates.

A formulated product containing sodium carbonate peroxohydrate with a synthetic, lamellar silicate having ion-exchange properties is sold by Hoechst (Germany) under the trade name SKS-9 PC. It is intended for incorporation in special washing and cleaning agents as well as for bleaching and disinfecting.

Sodium carbonate peroxohydrate has been used as an oxidizing agent in synthetic organic chemistry, at least on the laboratory scale (19). In nonaqueous systems it serves as a source of concentrated hydrogen peroxide and thus can be used to make peroxycarboxylic acids from carboxylic anhydrides, chlorides, or imidazolides. The peroxohydrate does not need high solubility to react (97). A suspension of sodium carbonate peroxohydrate in tetrahydrofuran, or dichloromethane presaturated with water, can be used. These heterogeneous reactions can be accelerated by ultrasonic radiation (98).

The compound was first made commercially in England during World War II as a substitute for sodium peroxoborate, which could no longer be made because of a shortage of borax. It offers some advantages over the peroxoborate and has grown to be an important industrial chemical.

The LD_{50} (rat, oral) of sodium carbonate peroxohydrate is 1034 mg/kg (63). The occupational exposure limit is 10 mg/m³ per 40-hour week. The compound is a skin and eye irritant; inhalation of dust can cause irritation to the mucous membranes and the respiratory system. It should be handled by using eye protection, rubber gloves, and industrial footwear. It is an oxidizing agent that can be decomposed by water, direct sources of heat, and catalysts. Decomposition is accompanied by liberation of oxygen and heat, which can support combustion and cause pressure bursts in confined spaces. Decomposition in the presence of organic material can be rapid and highly exothermic. The product is considered by the United Nations as nondangerous for transport purposes.

Sodium carbonate peroxohydrate is made by Chang Chun (Taiwan), Degussa (Germany), Eilenberg (Germany), Mitsubishi Gas Chemical (Japan), Oriental Chemical (Korea), Riverside Products (United States), Solvay Interlox (Germany, United Kingdom, United States), Tokai Denka, and Treibacher Chemische Werke. Trade names in use are OXYPER and FB Sodium Percarbonate. World production in 1993 was about 60,000 t; the U.S. price was \$1.41/kg, the U.K. price was £750/tonne (\$1.11/kg).

Other Peroxohydrates. Potassium, rubidium, and cesium carbonates all form peroxohydrates having the general formula $\text{M}_2\text{CO}_3 \cdot 3\text{H}_2\text{O}$. Crystal structures have not been established; Raman spectra (31) confirm the presence of molecular hydrogen peroxide in the crystal. These compounds are unstable and have no commercial application.

Ammonium carbonate peroxohydrate, $(\text{NH}_4)_2\text{CO}_3 \cdot \text{H}_2\text{O}_2$, first reported in 1980 (31), is crystallized from a solution of ammonium hydrogen carbonate in aqueous hydrogen peroxide. The vibrational spectrum confirms the presence of molecular hydrogen peroxide. The compound is unstable and unlikely to find commercial application.

Urea forms a 1:1 adduct with hydrogen peroxide. Urea peroxohydrate [124-43-6], $\text{CO}(\text{NH}_2)_2 \cdot \text{H}_2\text{O}_2$, is made simply by mixing powdered urea and 35% hydrogen peroxide in the presence of stabilizers, and crystallizing the product by cooling or concentration. It is available in the form of crystals or tablets. The former contain about 35% H_2O_2 , the latter about 34%. The solubility in water is 510 g/L at 20°C. The solution decomposes above 55°C.

Obsolete uses of urea peroxohydrate, as a convenient source of aqueous hydrogen peroxide, include the chemical deburring of metals, as a topical disinfectant and mouth wash, and as a hairdresser's bleach. In the 1990s the compound has been studied as a laboratory oxidant in organic chemistry (99,100). It effects epoxidation, the Baeyer-Villiger reaction, oxidation of aromatic amines to nitro compounds, and the conversion of sodium and nitrogen compounds to S—O and N—O compounds.

Urea peroxohydrate is an irritant to skin, eyes, and mucous membranes (2). The U.S. Food and Drug Administration approves it as an over-the-counter drug. It is slightly hygroscopic and sensitive to decomposition by catalytic materials such as heavy metals and their salts; it is also a flammable solid with oxidizing properties. Mixtures with combustible materials are easily ignited by friction or impact and burn vigorously. It has been assigned UN No. 1511 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances.

Urea peroxohydrate is made commercially by Solvay Deutschland, Degussa (Germany), and Mitsubishi Gas Chemical. It is known commercially as urea hydrogen peroxide, hydrogen peroxide carbamide, Exterol, Hydroperit, Hydroperit, Hyperol, Ortizon, Percarbamid, Percarbamide, Perhydrit, Perhydrol-Urea, Thenardol, and UHP. In 1994 the U.K. price was £7–8/kg (\$10–12/kg). World production in 1993 was several hundred metric t.

A related compound, the peroxohydrate of melamine, $\text{C}_3\text{H}_3\text{N}_3$, was made by Peroxid-Chemie but discontinued in the 1980s. In addition, a diperoxohydrate of the organic base DABCO is recognized and has been used as an oxidant in organic chemistry, but such a system has been known to explode (101).

Sodium pyrophosphate peroxohydrate, $\text{Na}_4\text{P}_2\text{O}_7 \cdot 2\text{H}_2\text{O}_2$, was made commercially by FMC in the 1960s but was discontinued. It is less stable than sodium peroxoborate and sodium carbonate peroxohydrate in the solid state, but more stable than these in solution. Many other phosphate peroxohydrates of the Group 1 and Group 2 metals are known but have found no uses.

Sodium sulfate peroxohydrate hydrate [29381-14-4], $2\text{Na}_2\text{SO}_4 \cdot \text{H}_2\text{O}_2 \cdot 2\text{H}_2\text{O}$, has been studied as a possible ingredient for domestic detergents because it conveniently combines two common detergent ingredients. A related compound containing sodium chloride, $4\text{Na}_2\text{SO}_4 \cdot 2\text{H}_2\text{O}_2 \cdot \text{NaCl}$, is more stable but the chloride ion is undesirable in detergents. Neither of these has been commercialized.

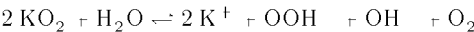
Peroxopolyoxometallates

Polyoxometallates, derived from both isopoly acids and heteropoly acids, are important homogeneous oxidation catalysts (102,103). The metals involved are vanadium, niobium, tantalum, molybdenum, and tungsten. The reactions involved are the oxidation of a wide range of organic compounds by hydrogen peroxide or organic hydroperoxide. The tungsten systems, first studied in the 1980s by Venturello (104) and Ishii (105), have come to be known as the Venturello-Ishii systems (106). Crystalline peroxometallates have been isolated from some of these systems. A typical example is tris[tetra-*n*-hexylammonium] peroxophosphotungstate, [N(C H₁₃)₄][PO₄[W(O)(O₂)₂]₄]. Whether such peroxopolymetallates are in fact intermediates in the reactions from which they have been isolated is a matter for research in each case. Solutions of peroxometallates, made by acidifying molybdate and tungstate solutions in the presence of hydrogen peroxide, have been studied by ir and Raman spectroscopy, and a large number of peroxo species have been detected (107,108).

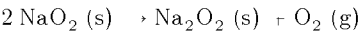
Superoxides

The superoxides are ionic solids containing the superoxide, O₂⁻. A comprehensive review of the superoxides was published in 1963 (109); they are described in Reference 1. Superoxides of all of the alkali metals have been prepared. Alkaline-earth metals, cadmium, and zinc all form superoxides, but these have been observed only in mixtures with the corresponding peroxides (110). The tendency to form superoxides in the alkali metal series increases with increasing size of the metal ion.

Metal superoxides are yellow-to-orange solids. Strong oxidizing agents, they react vigorously with most organic materials and reducing agents, and oxidize many metals to their highest oxidation states. One of the most prominent properties is the evolution of oxygen and hydroperoxide ion by reaction with water:



Sodium superoxide [12034-12-7], NaO₂, is a yellow solid, thermochemically unstable at ambient conditions with respect to the following reaction:



The compound has been prepared in reasonably high purity by the reaction of oxygen with sodium peroxide at 490°C and 298 atm (111). No applications are known.

Potassium superoxide [12030-88-5], KO₂, is a canary yellow solid that melts at 450–500°C when pure. It is paramagnetic and has a magnetic moment of 1.9 × 10⁻²³ J T⁻¹ (2.05 μ_B). Its heat of formation is ca 285 kJ mol⁻¹ (68 kcal mol⁻¹) and the specific heat is ca 79.5 J (mol·K)⁻¹ (19 cal (mol·C)⁻¹. The oxygen-dissociation pressures of KO₂ at various temperatures are as follow (to convert kPa to mm Hg, multiply by 7.5) (112):

Temperature, °C	198	279	461	539	589
O ₂ pressure, kPa	4.0	14.8	43.1	58.8	73.3

Potassium superoxide, a strong oxidizing agent, is similar to the Group 1 metal peroxides. This superoxide reacts vigorously with water, yielding oxygen as well as an alkaline peroxide solution that is susceptible to further decomposition. The mechanism of the reaction with water is not well understood. Potassium superoxide reacts with moist CO₂ below 10°C, yielding oxygen and potassium peroxocarbonate. Above 50°C, the peroxocarbonate decomposes to additional oxygen, potassium carbonate, and potassium hydrogen carbonate.

Potassium superoxide is produced commercially by spraying molten potassium into an air stream, which may be enriched with oxygen. Excess air is used to dissipate the heat of reaction and to maintain the temperature at ca 300°C. It can also be prepared in a highly pure state by oxidizing potassium metal that is dissolved in liquid ammonia at -50°C.

Mine Safety Appliances Co. (MSA) manufactures potassium superoxide in the United States for use in self-contained breathing equipment (see OXYGEN GENERATION SYSTEMS). In this equipment, the exhaled air is held within the device where it contacts a bed of KO₂. Moisture in the exhaled air leads to complete decomposition of the KO₂, and CO₂ is absorbed by the resulting KOH. This type of equipment is used for rescue and firefighting purposes as well as for space and undersea exploration. There are several published uses for potassium superoxide in organic chemistry, eg, for oxidizing aromatic compounds (113,114) and for initiating anionic polymerization (115).

On contact with skin and mucous membranes, potassium superoxide is converted to potassium hydroxide, which is corrosive and irritating. Protective clothing should be worn when handling it. The reaction with moisture is exothermic and may induce further decomposition with the production of oxygen. This product has been assigned UN No. 2466 and should be transported in accordance with international transport regulations pertaining to Class 5.1, oxidizing substances. MSA manufactures potassium superoxide for use in its own engineering systems, and L'Air Liquide manufactures it for general sale. In 1994, the price in France was Fr. 300 kg (\$34 kg).

Rubidium superoxide [12137-25-6], RbO₂, and cesium superoxide [12018-61-0], CsO₂, are formed by direct reaction of the elements, but are most readily prepared by oxidation of solutions of the metals in liquid ammonia. They are not produced commercially.

Calcium superoxide [12133-35-6], Ca(O₂)₂; strontium superoxide [12169-21-0], Sr(O₂)₂; and barium superoxide [55837-89-3], Ba(O₂)₂, have all been obtained in low yield and purity by treating the corresponding peroxides with hydrogen peroxide and heating (116). Heating the metal peroxides or their peroxohydrates often yields products containing some superoxides. These superoxides are not produced commercially.

Ozonides

The ozonides are characterized by the presence of the ozonide ion, O₃⁻. They are generally produced by the reaction of the inorganic oxide and ozone (qv). Two reviews of ozonide chemistry are available (1,117). Sodium ozonide [12058-54-7], NaO₃; potassium ozonide [12030-89-6], KO₃; rubidium ozonide [12060-04-7], RbO₃; and cesium ozonide [12053-67-7], CsO₃, have all been reported (1). Ammonium ozonide [12161-20-5], NH₄O₃, and tetramethylammonium ozonide [78657-29-1], (CH₃)₄N O₃, have been prepared at low temperatures (118).

The potassium salt is the best characterized. It is an orange-red paramagnetic solid having a magnetic moment of 1.6 × 10⁻²³ J T⁻¹ (1.73 μ_B). It reacts with water, yielding oxygen gas and potassium hydroxide. It decomposes to the superoxide, KO₂, upon standing at room temperature. Potassium ozonide is prepared by ozonation of dry potassium hydroxide. It can be purified by extraction and recrystallization from liquid ammonia. Whereas the inorganic ozonides are of potential importance as solid-oxygen carriers in breathing apparatus, they are not produced commercially.

Economic Aspects

All of the large-tonnage peroxo compounds, eg, sodium peroxoborate hexahydrate, sodium peroxoborate, and sodium carbonate peroxohydrate, are made by hydrogen peroxide producers using captive hydrogen peroxide. The world demand for active oxygen provided by these products is fairly stable, rising with the gross national product. The demand for each product fluctuates from year to year, however, with the requirements of the detergents industry. There is some limited scope, in the medium term, for the manufacturers to interchange their production facilities for the three products to cope with this

fluctuating demand.

In 1993, the world capacity for sodium peroxoborate hexahydrate (tetrahydrate) was about 900,000 metric tons, of which about one-third was converted to the dehydrated compound (monohydrate). At the then prevailing prices, the total value of this business was about $\$6 \times 10^9$. The world capacity for sodium carbonate peroxohydrate (percarbonate) was about 60,000 metric tons, valued at about $\$70 \times 10^6$.

Because the peroxodisulfate salts are all made electrochemically, the electrical energy cost is a significant part of their manufacturing cost. The 1994 world capacity for peroxodisulfate salts was about 75,000 metric tons, valued at about $\$30 \times 10^6$. The principal applications are in polymerization catalysis and the market broadly tracks the plastics business. The Caro's acid business is difficult to quantify because the product itself is not commercial but made on-site from purchased hydrogen peroxide.

Of the binary peroxides made from hydrogen peroxide, calcium peroxide is the most important. World production is about 2000 t yr, which is dominated by the dough-conditioning market in the United States. The markets for the other binary peroxides, such as those of zinc, magnesium, and strontium, total only a few hundred metric tons. Sodium peroxide and potassium superoxide are made from the alkali metals and their total markets are in the hundreds of tons.

BIBLIOGRAPHY

"Peroxides, Inorganic" in *ECT* 1st ed., Vol. 10, pp. 38–58, by E. S. Shanley, Buffalo Electrochemical Co., Inc., H. N. Gilbert and D. Raymond, E. I. du Pont de Nemours & Co., and J. Kleinberg, University of Kansas; "Peroxides and Peroxy Compounds, Inorganic" in *ECT* 2nd ed., Vol. 14, pp. 746–765, by E. S. Shanley, Arthur D. Little, Inc., and John O. Edwards, Brown University; in *ECT* 3rd ed., Vol. 17, pp. 1–26, by R. E. Hall, FMC Corp.

1. I. I. Vol'nov, *Peroxides, Superoxides, and Ozonides of Alkali and Alkaline Earth Metals*, Plenum Press, New York, 1966.
2. R. J. Lewis, Jr., ed., *Sax's Dangerous Properties of Industrial Materials*, Van Nostrand Reinhold, Co., Inc., New York, 1992.
3. P. Kleinschmidt, in W. Gerhartz, ed., *Ullman's Encyclopedia of Industrial Chemistry*, 5th ed., Vol. A19, VCH, Weinheim, Germany, 1991, p. 180.
4. A. E. Comyns, *Dictionary of Named Processes in Chemical Technology*, Oxford University Press, U.K., 1993.
5. Ref. 1, p. 36.
6. **Brit. Pat. 2,219,474** (May 5, 1989), D. B. Mobbs, S. P. Ball, and A. J. Monhemius (to Interlox Chemicals Ltd.).
7. R. E. Tieckelmann and R. E. Steele, *Food Tech.* **45**(1), 106 (1991).
8. **U.S. Pat. 1,385,842** (July 26, 1921), C. J. Patterson (to Campbell Baking Co.).
9. **U.S. Pat. 1,775,037** (Sept. 2, 1930), A. H. Fiske (to Rumford Chemical Works).
10. N. G. Vannerberg, *Arkiv Kemi.* **14**, 119 (1959).
11. A. Hansson, *Acta Chem. Scand.* **15**, 934 (1961).
12. M. A. A. F. de C. T. Carrondo and A. C. Skapski, *Acta Cryst.* **B34**, 3551 (1978).
13. C. A. Morgan, *Mellor's Comprehensive Treatise on Inorganic and Theoretical Chemistry*, Suppl., Vol. 5. Longman, London, 1980.
14. J. Flanagan and co-workers, *J. Chem. Soc. Dalton Trans.*, 1651 (1989).
15. R. Pizer and C. Tihal, *Inorg. Chem.* **26**, 3639 (1987).
16. B. N. Chernyshov and co-workers, *Koord. Khimya*, **11**(1), 31 (1985).
17. B. N. Chernyshov and O. V. Brovkina, *Latv. PSR Zinat. Akad. Vestis Kim. Ser.* (4), 422 (1988).
18. W. P. Griffith and co-workers, *J. Chem. Soc. Faraday Trans.* **89**, 4035 (1993).
19. A. McKillop and W. R. Sanderson, *Tetrahedron* **51**(22), 6145 (1995).
20. A. McKillop and J. A. Tarbin, *Tetrahedron Letters*, **24**(14), 1508 (1983).
21. A. McKillop and J. A. Tarbin, *Tetrahedron*, **43**(8), 1753 (1987).
22. C. Karaunakaran, K. Ganapathy, and P. Ratnasamy, *Reac. Kinet. Catal. Lett.* **40**(2), 369 (1989).
23. K. N. Reddy, K. C. Rajanna, and P. K. Saiprakash, *Orient. J. Chem.* **4**(2), 159 (1988).
24. C. Karunakaran and P. Manimekalai, *Tetrahedron*, **47**(41), 8733 (1991).
25. R. H. Dreisbach, *Handbook of Poisoning*, 12th ed., Lange Medical Publications, Los Altos, Calif., 1987, p. 360.
26. W. P. Griffith, A. C. Skapski, and A. P. West, *Chem. Ind.*, 185 (1984).
27. J. Flanagan and co-workers, *Spectrochim. Acta.*, **45A**(9), 951 (1989).
28. *Registry of Toxic Effects of Chemical Substances* (RTECS), NIOSH, Washington, D.C., Sept. 1994.
29. R. Bruce and co-workers, *J. Am. Chem. Soc.* **87**, 2057 (1965).
30. J. O. Edwards and co-workers, *J. Am. Chem. Soc.* **91**, 1095 (1969).
31. D. P. Jones and W. P. Griffith, *J. Chem. Soc. Dalton Trans.*, 2526 (1980).
32. A. Blaschette, *Z. Anorg. Chem.* **384**, 177 (1971).
33. A. Blaschette and S. M. Saad, *Holzforschung*, **23**, 163 (1969).
34. W. P. Griffith, private communication, London, 1994.
35. G. Rietz and H. Kopp, *Z. Anorg. Chem.* **382**, 31 (1971).
36. G. Rietz and J. Loescher, *Z. Anorg. Chem.* **382**, 37 (1971).
37. G. Rietz and H. Kopp, *Z. Anorg. Chem.* **384**, 19 (1971).
38. *Gmelin's Handbook of Inorganic and Organometallic Chemistry*, System No. 46, Vol. C1, Springer-Verlag, Berlin, 1972, p. 139.
39. A. G. Davies and P. J. Smith, in G. Wilkinson, ed., *Comprehensive Organometallic Chemistry*, Vol. 11, Pergamon Press, Oxford, U.K., 1982, p. 581.
40. W. G. Keith and R. E. Powell, *J. Chem. Soc. A*, 90 (1969).
41. D. J. Benton and P. Moore, *J. Chem. Soc. A*, 3179 (1970).
42. F. Raschig, *Z. Angew. Chem.* **17**, 1419 (1904).
43. J. O. Edwards and R. C. Plumb, *Prog. Inorg. Chem.* **41**, 599 (1994).
44. C. J. Battaglia and J. O. Edwards, *Inorg. Chem.* **4**, 552 (1965).
45. C. Toennies, *J. Am. Chem. Soc.* **59**, 555 (1937).
46. A. Simon and H. Richter, *Z. Anorg. Allg. Chem.* **304**, 1 (1960).
47. M. M. Crutchfield and J. O. Edwards, *J. Am. Chem. Soc.* **82**, 3533 (1960).
48. M. M. Crutchfeld, in J. O. Edwards, ed., *Peroxide Reaction Mechanisms*, Wiley-Interscience, New York, 1962.
49. F. Fichter and W. Bladergroen, *Helv. Chim. Acta.* **9**, 559 (1927).
50. W. P. Griffith, R. D. Powell, and A. C. Skapski, *Polyhedron*, **7**(14), 1305 (1988).
51. I. N. Vol'nov, U. N. Chamova, and E. I. Kolotova, *Russ. J. Inorg. Chem.* **6**(2), 137 (1961).
52. **U.S. Pat. 3,616,325** (Oct. 26, 1971), P. R. Mucenieks (to FMC Corp.).
53. **U.S. Pat. 3,634,024** (Jan. 11, 1972), R. E. Yellin and R. F. Villiers (to FMC Corp.); **U.S. Pat. 3,649,164** (Mar. 14, 1972), R. E. Yellin, L. A. Sitver, and R. F. Villiers (to FMC Corp.).
54. *Textile Chem. Color.* **5**(5), 79 (1973).
55. E. C. Moreno, M. Kresak, and A. Gaffar, *J. Colloid, Interface Sci.* **168**(1), 173 (1994).
56. **Brit. Pat. 1,585,011** (Feb. 18, 1981), J. P. Schmirmann, M. Pralus, and S. Y. Delavarenne (to Produits Chimiques Ugine Kuhlmann); **U.S. Pat. 3,993,673** (Nov. 23, 1976), C. H. Mullen (to Union Carbide Corp.).
57. S. E. Jacobsen, F. Mares, and P. M. Zambri, *J. Am. Chem. Soc.* **101**, 6946 (1979).

58. M. Spiro, *Electrochim. Acta*, **24**, 313 (1979).

59. W. V. Steele and E. H. Appelman, *J. Chem. Thermo.* **14**, 337 (1982).

60. **Ger. Pat. 2,336,996** (Jan. 31, 1974), J. P. Zumbunn (to L'Air Liquide SA); **Fr. Pat. 2,163,958** (Aug. 31, 1973), J. P. Zumbunn (to L'Air Liquide SA).

61. R. R. Lastomirsky, *Diss. Abstr. B.* **30**, 3098 (1970).

62. M. Spiro, private communication, London, 1995.

63. P. J. Wyborn, private communication, Widnes, 1994.

64. C. Declerck, private communication, Brussels, 1977.

65. **U.S. Pat. 5,304,360** (Apr. 19, 1994), J. R. G. Lane, C. F. McDonagh, and S. E. Woods (Interox Chemicals Ltd.).

66. E. N. Wilton, P. J. Wyborn, and J. A. Reeve, *Proceedings of the 1st European Metals Conference, Brussels, 1991*, Elsevier Applied Science, New York, 1991, p. 423.

67. H. M. Castrantas and co-workers, *Randol Gold Forum*, Randol International Ltd., Golden, Colo., 1993.

68. N. A. Troughton, F. Desprez, and J. Devenyns, paper presented to the 1994 Non-Chlorine Bleaching Conference, Amelia Island, Fla., Mar. 1994.

69. J. Flanagan, W. P. Griffith, and A. C. Skapski, *J. Chem. Soc. Chem. Comms.*, 1574 (1984).

70. W. Franck and B. Bertsch-Frank, in Ref. 3, p. 188.

71. R. W. Murray, *Chem. Revs.* **89**, 1187 (1989).

72. W. Adam, R. Curci, and J. O. Edwards, *Accts. Chem. Res.* **22**, 205 (1989).

73. H. Marshall, *J. Chem. Soc.* **59**, 771 (1891).

74. W. H. Zachariasen and R. C. L. Mooney, *Z. Kristallogr.* **88**, 63 (1934).

75. C. A. Crampton and co-workers, in R. Thompson, ed., *The Modern Inorganic Chemicals Industry*, The Chemical Society, London, 1977, p. 259.

76. **Ger. Pat. 2,346,945** (Mar. 27, 1975), E. Rossberger (to Peroxid-Chemie GmbH).

77. L. J. Csanyi, *Naturwiss.* **43**, 33 (1956).

78. D. D. Wagman and co-workers, *U.S. National Bureau of Standards, Technical Note 270* (1981).

79. D. A. House, *Chem. Revs.* **62**, 185 (1962).

80. K. Ogawa and Y. Nomura, *Yuki Gosei Kagaku Kyokai-shi*, **42**(2), 98 (1984).

81. H. Wurziger, *Kontakte (Darmstadt)* (3), 30 (1987).

82. J. A. Connor and E. A. V. Ebsworth, *Adv. Inorg. Chem. & Radiochem.* **6**, 279 (1964).

83. K. A. Jorgensen, *Chem. Rev.* **89**(3), 431 (1989).

84. N. Kitajima, M. Akita, and Y. Moro-oka, in W. Ando, ed., *Organic Peroxides*, John Wiley & Sons, Ltd., Chichester, U.K., 1992, p. 535.

85. V. Conte, F. di Furia, and G. Modena, in Ref. 84, p. 559.

86. G. Strukul, in G. Strukul, ed., *Catalytic Oxidations with Hydrogen Peroxide as Oxidant*, Kluwer Academic Publishers, Dordrecht, 1993, Chapt. 6.

87. R. A. Sheldon and J. K. Kochi, *Metal-Catalysed Oxidations of Organic Compounds*, Academic Press, Inc., New York, 1981.

88. H. Mimoun, in Wilkinson, ed., *Comprehensive Coordination Chemistry*, Vol. 6, Pergamon Press, Oxford, U.K., 1987, p. 317.

89. C. Keller, *The Chemistry of the Actinides*, Pergamon Press, Oxford, U.K., 1973, p. 248.

90. I. Wells and D. Pepper, in J. F. Flagg, ed., *Chemical Processing of Reactor Fuels*, Academic Press, Inc., New York, 1961, p. 339.

91. J. K. Dawson and R. G. Sowden, *Chemical Aspects of Nuclear Reactors*, Butterworths, London, 1963, p. 90.

92. W. C. Schumb, C. N. Satterfield, and R. L. Wentworth, *Hydrogen Peroxide*, Reinhold Publishing Corp., New York, 1955, pp. 26, 641.

93. Ref. 75, p. 256.

94. M. A. A. F. de C. T. Carrondo and co-workers, *J. Chem. Soc. Dalton Trans.*, 2323 (1977).

95. J. Flanagan and co-workers, *J. Chem. Soc. Chem. Comms.*, 20 (1986).

96. **U.S. Pat. 4,526,698** (July 2, 1985), M. Kurada, M. Mutata, and T. Tamura (to Kao Corp.).

97. A. M. d'A. Roche-Gonsalves and co-workers, *J. Chem. Res. (S)*, 208 (1991).

98. X. Gao-Yang and co-workers, *Acta Chim. Sin.* (Eng. ed.), (5), 463 (1989).

99. H. Heany, *Aldrichimica Acta*, **26**(2), 35 (1993).

100. M. S. Cooper and co-workers, *Synlett*, 553 (1990).

101. P. G. Cookson, A. G. Davies, and N. Fazal, *J. Organometal. Chem.* **99**, C31 (1975).

102. W. P. Griffith, *Trans. Met. Chem.* **16**, 548 (1991).

103. C. L. Hill, in Ref. 86, Chapt. 8.

104. C. Venturello, E. Alneri, and M. Ricci, *J. Org. Chem.* **48**, 3831 (1983).

105. Y. Ishii and co-workers, *J. Org. Chem.* **53**, 3587 (1988).

106. C. L. Hill and co-workers, in D. H. R. Barton, A. E. Martell, and D. T. Sawyer, eds., *The Activation of Dioxygen and Homogeneous Catalytic Oxidation*, Plenum Press, New York, 1993, p. 71.

107. N. J. Campbell and co-workers, *J. Chem. Soc. Dalton Trans.* 1203 (1989).

108. A. C. Dengel, W. P. Griffith, and B. C. Parkin, *J. Chem. Soc. Dalton Trans.*, 2683 (1993).

109. A. W. Petrocelli, *J. Chem. Educ.* **40**, 146 (1963).

110. J. K. Thompson and J. Kleinberg, *J. Am. Chem. Soc.* **73**, 1243 (1951).

111. S. E. Stephanou and co-workers, *J. Am. Chem. Soc.* **71**, 1819 (1949).

112. R. A. Sheldon and J. A. Van Doorn, *J. Organomet. Chem.* **94**, 115 (1975).

113. C. Sotiriou, G. Wenni, and R. W. Giese, *J. Org. Chem.* **55**(7), 2159 (1990).

114. S. Abdel-Baky, C. Sotiriou, and R. W. Giese, *Tetrahedron* **47**(30), 5667 (1991).

115. Y. K. Han, J. H. Kim, and S. K. Choi, *J. Polym. Sci. Chem. Ed.*, **29**(2), 281 (1991).

116. T. P. Whaley and J. Kleinberg, *J. Am. Chem. Soc.* **73**, 79 (1951).

117. A. W. Petrocelli and R. V. Chiarenzelli, *J. Chem. Educ.* **39**, 557 (1962).

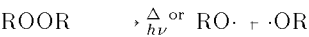
118. I. Solomon and co-workers, *J. Am. Chem. Soc.* **84**, 34 (1962).

Alan E. Comyns
Solvay Interox

ORGANIC PEROXIDES

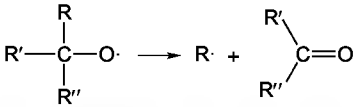
Organic peroxides are compounds possessing one or more oxygen–oxygen bonds. They are derivatives of hydrogen peroxide, HOOH, in which one or both hydrogens are replaced by a group containing carbon (R, R'), ie, ROOH or ROO'. The ultimate source of the oxygen–oxygen linkage in organic peroxides is oxygen; either from direct air oxidation or from reactions of organic compounds with peroxidic materials derived from oxygen, eg, hydrogen peroxide, alkali metal peroxides, ozone (qv), or other organic peroxides (1–6). Organic peroxides are intermediates or products in air oxidation of many synthetic and natural organic compounds. They are involved in many biological processes including development of rancidity in fats, loss of activity of vitamin products, and firefly bioluminescence. Some biological products contain a peroxide group, eg, the natural product, qinghaosu [63968-64-9], is a 1,2,4-trioxane that possesses antimalarial properties (7) and ascaridole [512-85-6], is an endoperoxide that possesses sedative, analgesic, antirheumatic, and anthelmintic properties (8). Organic peroxides are also involved in gum formation in lubricating oils, prepolymerization of some vinyl monomers, and degradation of olefin polymers. Air oxidation of certain solvents, especially ethers, results in formation of organic peroxides which, upon concentration, can form highly explosive peroxidic residues. Oxidation inhibitors are often used to prevent formation of undesirable peroxides in products.

Almost all organic peroxides are thermally and photolytically sensitive owing to the facile cleavage of the weak oxygen–oxygen bond, ie, the range of Δ*H* is about 84 to 184 kJ mol⁻¹ (20 to 44 kcal mol⁻¹) (9–11):

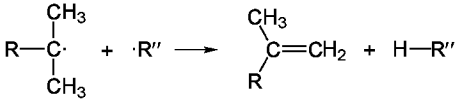


This cleavage is a unimolecular (first-order) reaction. The thermal decomposition rates are affected by the structure of the organic peroxide and the decomposition conditions. For comparison purposes, the thermal activity of a peroxide can be expressed in terms of its 10-hour half-life temperature (10-h HLT, ie, the temperature required for 50% decomposition of a peroxide in a period of 10 hours). This temperature can vary from well below 20°C to well above 200°C. In comparing two peroxides, the one with the lower 10-h HLT is more thermally labile. Another comparison of thermal activity is by active oxygen content, which refers to the quantity of peroxide groups available for thermal decomposition. The concept of active oxygen content is based on the existence of one active oxygen in each oxygen–oxygen bond pair. It is usually expressed as a percentage, ie, (100)/(16 *p m*), where *p* represents the number of peroxide groups present and *m* represents the molecular weight (pure compound) and is adjusted for assay in diluted formulations. The phrase appears frequently in discussions of peroxide analysis and safety. Peroxide compounds or formulations having low active oxygen content are generally safer to handle than those with high active oxygen content.

Thermal decomposition of peroxides initially forms oxygen-centered free radicals from the oxygen–oxygen bond homolysis. These radicals are reactive intermediates generally having very short lifetimes, ie, half-life times less than 10⁻³ s (12). Once formed, radicals undergo two basic types of reactions: propagating reactions and terminating reactions. In a propagating reaction, a radical reacts to form a covalent bond and to generate a new radical. The three most common propagating reactions are hydrogen abstraction (eq. 1), β-scission (eq. 2), and addition to carbon–carbon double bonds or aromatic rings (eq. 3).



Generally, the higher stability of the generated alkyl radical compared to that of the starting radical provides the driving force and determines the course of a propagating reaction. The radical, R·, formed in the β-scission reaction is the most stable alkyl radical among the three possible radicals, R·, R'·, and R''·. Propagating reactions are also affected by temperature, pressure, steric, and electronic effects. In a termination reaction, two radicals interact in a mutually destructive reaction in which both radicals form covalent bonds and reaction ceases. The two most common termination reactions are coupling (eq. 4) and disproportionation (eq. 5).



Because many organic peroxides undergo thermolysis to form useful free radicals, they are used commercially as initiators for free-radical reactions. Many organic peroxides also undergo reactions in which free radicals are not involved, eg, heterolyses, hydrolyses, reductions, and rearrangements. Numerous reviews of the chemistry and applications of organic peroxides have been published (11,13–41).

The first synthesis of an organic peroxide was that of dibenzoyl peroxide [94-36-0] (BPO) in 1858 (42). In the early 1900s, BPO was employed as a bleaching agent for edible oils and, subsequently, for grain flours. The use of organic peroxides as free-radical initiators for vinyl monomer polymerization became important during World War II due to increased demand for plastics and synthetic rubber for military tires. Approximately 100 different organic peroxides in well over 300 formulations are commercially produced throughout the world as free-radical initiators for polymerizing vinyl monomers, grafting of monomers onto polymers, curing agents for unsaturated resins, rubber, and elastomers, cross-linking of thermoplastics (eg, polyethylene), modification degradation of polypropylene, halogenations, anti-Markovnikov additions to terminal olefins (eg, formation of primary mercaptans), and telomerizations. Some are used as bleaching agents (qv) (ie, for grain flours and fabrics), olefin epoxidizing agents, and active species in a variety of other applications, eg, the use of BPO as the active antibacterial component in acne medications.

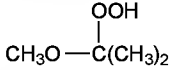
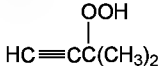
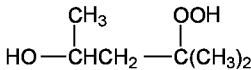
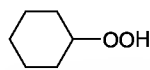
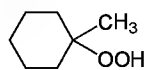
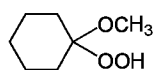
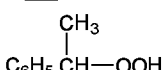
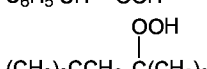
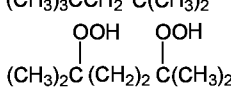
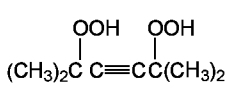
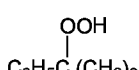
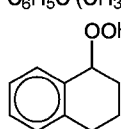
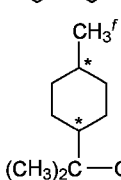
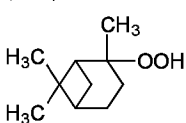
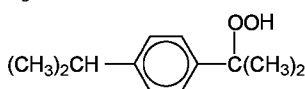
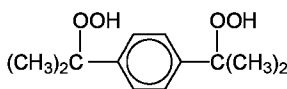
Organic peroxides can be classified according to peroxide structure. There are seven principal classes: hydroperoxides; dialkyl peroxides; α-oxygen substituted alkyl hydroperoxides and dialkyl peroxides; primary and secondary ozonides; peroxyacids; diacyl peroxides (acyl and organosulfonyl peroxides); and alkyl peroxyesters (peroxycarboxylates, peroxyulfonates, and peroxyphosphates).

Hydroperoxides

There are two main subclasses of hydroperoxides: organic (alkyl) hydroperoxides, ie, ROOH, and organomineral hydroperoxides, ie, R_m Q(OOH)_n, where Q is silicon (43), germanium, tin, or antimony. The alkyl group in ROOH can be primary, secondary, or tertiary. Except for ethylbenzene hydroperoxide, only *tert*-alkyl hydroperoxides are commercially important.

Physical Properties. Some physical properties of alkyl hydroperoxides (in order of increasing carbon content) are listed in Table 1 (44). Descriptions of hydroperoxides are given in the chemical literature (1,4–6,10,28,43,45).

Table 1. Properties of Some Alkyl Hydroperoxides^a

Hydro-peroxide	CAS Registry Number	Structure	Bp, °C (kPa) ^b	Mp, °C	<i>n</i> _D ²⁰
methyl	[3031-73-0]	CH ₃ –OOH	45.5–46.5 (24.53)		1.3654 ^c
ethyl	[3031-74-1]	C ₂ H ₅ –OOH	43–44 (6.67)		
isopropyl	[3031-75-2]	(CH ₃) ₂ CH–OOH	38–38.5 (2.67)		
<i>n</i> -butyl	[4813-50-7]	<i>n</i> -C ₄ H ₉ –OOH	40–42 (1.07)		1.4057
<i>sec</i> -butyl	[13020-06-9]	<i>sec</i> -C ₄ H ₉ –OOH	41–42 (1.47)		1.4050
<i>tert</i> -butyl	[75-91-2]	<i>t</i> -C ₄ H ₉ –OOH	33–42 (2.27)	4.0–4.5	1.3983 ^d
2-methoxy-2-propyl	[10027-74-4]		61–63 (2.40)		
<i>tert</i> -amyl	[3425-61-4]	<i>t</i> -C ₅ H ₁₁ –OOH	34–35 (0.93)		1.4120 ^d
1,1-di-methylpropynyl	[36566-81-1]		42 (2.27)		
3-hydroxy-1,1-dimethylbutyl	[66734-30-3]				1.4418 ^e
cyclohexyl	[766-07-4]		57 (0.16)	20	1.4622
<i>n</i> -heptyl	[764-81-8]	<i>n</i> -C ₇ H ₁₅ –OOH	42–43 (0.008)	35.5	1.4269
3-ethyl-3-pentyl	[18428-34-7]	(C ₂ H ₅) ₃ C–OOH	71–73 (2.27)	2–3	1.4379
1-methyl-cyclohexyl	[4952-03-8]		38 (0.004)		1.4652
1-methoxy-cyclohexyl	[16580-35-1]		54.5–55 (0.027)		
ethylbenzene	[3071-32-7]		48.2 (0.027)		1.5265
1,1,3,3-tetra-methyl-butyl	[5809-08-5]		44–45 (0.12)		
2,5-di-methyl-2,5-di-hydro-peroxy-hexane	[3025-88-5]			105	
2,5-di-methyl-2,5-dihydro-peroxy-3-hexyne	[3491-36-9]			107–109	
α-cumyl	[80-15-9]		60 (0.027)		1.5242
1,2,3,4-tetra-hydro-naph-thalene	[26447-24-5]		120–125 (0.027)	56	
<i>p</i> -men-thane	[26762-92-5]				1.4558 ^g
pinane	[28324-52-9]		57 (0.013)		
<i>p</i> -diisopro-pylbenzene	[98-49-7]			33–34	1.5134
mono-hydro-peroxide					
<i>p</i> -diisopro-pylbenzene	[3159-98-6]			140–141.5	
dihydro-peroxide					
stearyl	[56537-19-0]	<i>n</i> -C ₁₈ H ₃₇ –OOH		49–50	

^a Refs. 10 and 44.

^b To convert kPa to mm Hg, multiply by 7.5.
^c At 21°C.
^d At 25°C.
^e 94° assay material.
Courtesy of Elf Atochem North America, Inc.
^f The OOH group may alternatively be at the positions marked by an asterisk.
^g At 25°C for 54° *p*-menthane in *p*-menthane.

Alkyl hydroperoxides can be liquids or solids. Those having low molecular weight are soluble in water and are explosive in the pure state. As the molecular weight increases, ie, as the active oxygen content is reduced, water solubility and the violence of decomposition decrease. Alkyl hydroperoxides are stronger acids than the corresponding alcohols and have acidities similar to those of phenols. *tert*-Alkyl hydroperoxides can be purified through their alkali metal salts (28).

Bond dissociation energies (BDEs) for the oxygen–oxygen and oxygen– hydrogen bonds are 167–184 kJ mol (40.0–44.0 kcal mol) and 375 kJ mol (89.6 kcal mol), respectively (10,45). Heats of formation, entropies, and heat capacities of hydroperoxides have been summarized (9). Hydroperoxides exist as hydrogen-bonded dimers in nonpolar solvents and readily form hydrogen-bonded associations with ethers, alcohols, amines, ketones, sulfoxides, and carboxylic acids (46). Other physical properties of hydroperoxides have been reported (46).

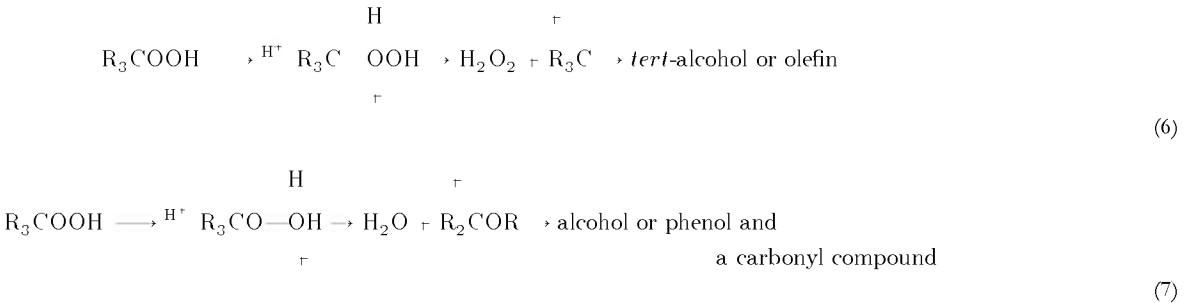
Chemical Properties. Hydroperoxides can react with or without cleavage of the oxygen–oxygen bond. Reactions resulting in scission of the oxygen–oxygen bond involve heterolytic, homolytic, or metal-promoted oxidation–reduction reactions. Alkyl hydroperoxides also react with a variety of compounds, eg, acyl chlorides, anhydrides, alcohols, alkyl halides, sulfates or sulfonates, epoxides, ethers, olefins, acetals, aldehydes, ketones, ketenes, phenols, amines, aralkanes, nitrites, isocyanates, chloroformates, phosgene, carbamyl chlorides, sulfonyl chlorides, and organomineral halides, amines, hydroxides, and ethers to form other organic peroxides in which the oxygen–oxygen bond is retained (47).

Alkyl hydroperoxides are reduced readily to the corresponding alcohols; many such reductions are quantitative and useful for analytical methods. Alkyl hydroperoxides have been used as oxidizing or hydroxylating reagents in organic syntheses, eg, to produce unsaturated acids and esters from unsaturated aldehydes; hydroxylamines, imines, and nitrones from amines; carboxylic acids from ketones; phosphates from phosphites; phosphine oxides from phosphines; sulfoxides from sulfides; alcohols and phenols from Grignard reagents; epoxides from olefins; phenols from aromatic compounds; ketones from secondary alcohols; stilbenequinones from *p*-methylphenols; and unsaturated ethers and hydroxyethers from dienes (1,4–6,10,28,32,48). A method has been developed for the asymmetric synthesis of optically active epoxy alcohols from unsaturated alcohols using hydroperoxides and tetraalkyl titanates as the epoxidizing system, and optically active tartrate as the asymmetric template (49).

Bases, such as potassium or sodium hydroxide, piperidine, and pyridine, react with primary and secondary hydroperoxides to form aldehydes or ketones (28). In some cases, this reaction is slow or fails unless heating is employed.

tert-Alkyl hydroperoxides form stable alkali metal salts with caustic; however, when equimolar amounts of the hydroperoxide and its sodium salt are present in aqueous solution, rapid decomposition to *tert*-alcohol and oxygen occurs (28).

Acids react with alkyl hydroperoxides in two different ways, depending on the hydroperoxide structure and the acid strength (45).



Carbon–oxygen cleavage occurs when protonation takes place on the oxygen atom adjacent to the carbon (eq. 6). Thus hydrogen peroxide is the leaving group and the resulting carbonium ion produces *tert*-alcohol or olefin. The more likely oxygen–oxygen bond cleavage occurs when protonation takes place on the second oxygen (eq. 7). In this case the oxygen atom adjacent to the carbon possesses a partial positive charge and one alkyl (or aryl) group migrates from the carbon to the adjacent positive oxygen via a 1,2-shift. Reaction 7 is especially rapid with aralkyl hydroperoxides. The commercial process for production of phenol (qv) and acetone (qv) involves the acid-catalyzed decomposition of α -cumyl hydroperoxide according to this reaction. Strong acids can also decompose hydroperoxides at ca 20°C by oxygen–oxygen bond homolysis. This homolysis accounts for about 30° of the decomposed hydroperoxide and can be used to initiate free-radical polymerization (50). Addition of strong acids, eg, concentrated sulfuric acid, to pure or concentrated solutions of alkyl hydroperoxides is dangerous and, unless carefully controlled, results in explosive decompositions and fire.

Hydroperoxides are photo- and thermally sensitive and undergo initial oxygen–oxygen bond homolysis, and they are readily attacked by free radicals undergoing induced decompositions (eqs. 8–10).



Therefore, first-order, decomposition rates for alkyl hydroperoxides, ie, from oxygen–oxygen bond homolysis, are valid only if induced decomposition reactions are suppressed completely. This is accomplished only if the decomposition occurs in very dilute solutions and efficient free-radical scavengers are present. Radical-induced decompositions of *tert*-alkyl hydroperoxides are chain reactions (51,52). The alkoxy radicals from equations 9 or 10 react with undecomposed hydroperoxide to generate more alkoxy radicals according to equations 8 and 10. For example, thermal decomposition of *tert*-butyl hydroperoxide, which is either undiluted or in chlorobenzene solution at 140°C, or photolytic decomposition at lower temperatures, yields *tert*-butyl alcohol and oxygen in nearly quantitative amounts (52).

Although primary and secondary alkyl hydroperoxides are attacked by free radicals, as in equations 8 and 9, such reactions are not chain scission reactions since the alkylperoxy radicals terminate by disproportionation without forming the new radicals needed to continue the chain (53). Overall decomposition rates are faster than the true first-order rates if radical-induced decompositions are not suppressed.

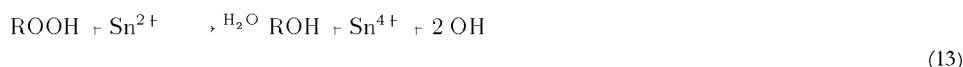
Hydroperoxides are decomposed readily by multivalent metal ions, ie, Cu, Co, Fe, V, Mn, Sn, Pb, etc, by an oxidation-reduction or electron-transfer process. Depending on the metal and its valence state, metallic cations either donate or accept electrons when reacting with hydroperoxides (45). Either one

or two electrons may be transferred depending on the metal:

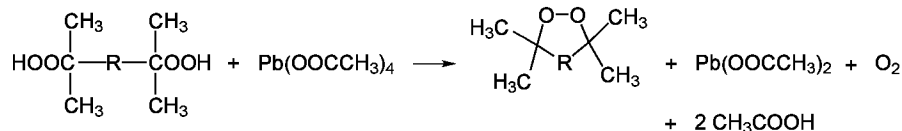


With most transition metals, eg, Cu, Co, and Mn, both valence states react with hydroperoxides via one electron transfer (eqs. 11 and 12). Thus, a small amount of transition-metal ion can decompose a large amount of hydroperoxide and, consequently, inadvertent contamination of hydroperoxides with traces of transition-metal impurities should be avoided.

An example of a reaction involving transfer of two electrons from the metal is the reduction of alkyl hydroperoxides with stannous chloride (10) (eq. 13).



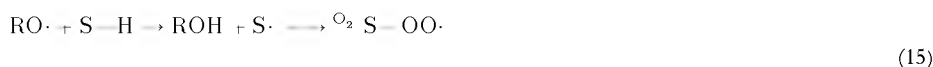
In the reaction of lead tetraacetate with 1,3- or 1,4-dihydroperoxides (10) to produce cyclic monoperoxides there are two electron transfers to the metal (eq. 14).



The reactions of *tert*-alkyl hydroperoxides with ferrous ion (eq. 11) generate alkoxy radicals. These free-radical initiator systems are used industrially for the emulsion polymerization and copolymerization of vinyl monomers, eg, butadiene–styrene. The use of hydroperoxides in the presence of transition-metal ions to synthesize a large variety of products has been reviewed (48,51).

The ultimate fate of the oxygen-centered radicals generated from alkyl hydroperoxides depends on the decomposition environment. In vinyl monomers, hydroperoxides can be used as efficient sources of free radicals because vinyl monomers generally are efficient radical scavengers which effectively suppress induced decomposition. When induced decomposition occurs, the hydroperoxide is decomposed with no net increase of radicals in the system (see eqs. 8, 9, and 10). Hydroperoxides usually are not effective free-radical initiators since radical-induced decompositions significantly decrease the efficiency of radical generation. Thermal decomposition-rate studies in dilute solutions show that alkyl hydroperoxides have 10-h HLTs of 133–172°C. Alkyl hydroperoxides are among the most thermally stable organic peroxides. However, hydroperoxides are sensitive to chain decomposition reactions initiated by radicals and/or transition-metal ions. Such decompositions, if not controlled, can be autoaccelerating and sometimes can lead to violent decompositions when neat hydroperoxides or concentrated solutions of hydroperoxides are involved.

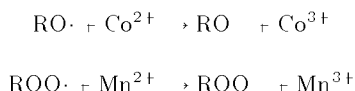
Most solvents for hydroperoxides are not completely inert to radical attack and, consequently, react with radicals from the hydroperoxide to form solvent-derived radicals, either by addition to unsaturated sites or by hydrogen- or chlorine-atom abstraction. In equation 15, S–H represents solvent and S· is a solvent radical.



Such solvent-derived radicals can induce the decomposition of the hydroperoxide or react with oxygen in the system to form peroxidic solvent molecules. They may also react with other radicals either by coupling or disproportionation.

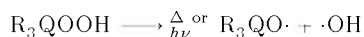
tert-Alkoxy radicals from hydroperoxides can undergo a β-scission reaction (eq. 2) to yield an alkyl radical and a ketone. The higher stability of the generated alkyl radical compared to that of the parent *tert*-alkoxy radical provides the driving force for this reaction, and the R group involved is the one that forms the most stable alkyl radical.

Transition-metal ions also interact with hydroperoxide-generated radicals by converting them into ions, eg:



The radicals are destroyed and are not available to take part in the desired radical reactions, eg, polymerizations. Thus, transition-metal ion concentrations of metal–hydroperoxide initiating systems are optimized to maximize radical generation.

Organomineral hydroperoxides undergo thermal and photolytic homolyses:



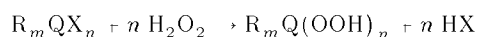
The decomposition products show that the initially formed radicals react by several mechanisms. These hydroperoxides are decomposed by bases, apparently by an ionic pathway, and are surprisingly stable in acids (33).

Synthesis. Hydroperoxides have been prepared from several types of peroxygen compounds including hydrogen peroxide or sodium peroxide, ozone, oxygen, and other organic peroxides (45). Hydrogen peroxide (H₂O₂) and its anions are powerful nucleophiles and react with reagents RX to form ROOH and HX, where X can be sulfate, acid sulfate, alkane- and arenesulfonate, chloride, bromide, hydroxyl, alkoxide, perchlorate, etc. RX can also be an alkyl orthoformate or *tert*-alkyl carboxylate.

Electron-rich olefins react with hydrogen peroxide under acidic conditions to form hydroperoxides, presumably by means of a carbonium ion intermediate, eg, *tert*-butyl hydroperoxide from isobutylene. *tert*-Butyl hydroperoxide has been produced commercially by mixing either *tert*-butyl alcohol or isobutylene with sulfuric acid followed by reaction with hydrogen peroxide. It also can be prepared by adding *tert*-butyl alcohol to peroxysulfuric acid [7722-86-3], but this method can be dangerous because an explosive composition is formed during addition (54). Other hydroperoxides produced commercially from hydrogen peroxide are *tert*-amyl hydroperoxide, 1,1,3,3-tetramethylbutyl hydroperoxide, 3-hydroxy-1,1-dimethylbutyl hydroperoxide, 2,5-dimethyl-2,5-dihydroperoxy-3-hexyne, and 2,5-dimethyl-2,5-dihydroperoxyhexane.

In the preparation of hydroperoxides from hydrogen peroxide, dialkyl peroxides usually form as by-products from the alkylation of the hydroperoxide in the reaction mixture. The reactivity of the substrate (olefin or RX) with hydrogen peroxide is the principal restriction in the process. If elevated temperatures or strongly acidic or strongly basic conditions are required, extensive decomposition of the hydrogen peroxide and the hydroperoxide can occur.

Organomineral hydroperoxides have been prepared from hydrogen peroxide and organomineral halides, hydroxides, oxides, peroxides, and amines (10,33). If HX is an acid, ammonia is used to prevent acidic decomposition.



Many hydroperoxides have been prepared by autoxidation of suitable substrates with molecular oxygen (45,52,55). These reactions can be free-radical chain or nonchain processes, depending on whether triplet or singlet oxygen is involved. The free-radical process consists of three stages:

initiation, propagation, and termination. The reaction can be initiated by free-radical initiators, eg, peroxides, aliphatic azo compounds, multivalent metal ions, uv radiation, etc (see INITIATORS). However, with many organic substrates, this is not required because oxygen serves as an initiator, probably by means of hydrogen abstraction at elevated temperatures: $\text{RH} + \text{O}_2 \rightarrow \text{R}\cdot + \text{HOO}\cdot$. Such autoxidations usually are characterized by induction periods. In the propagation reactions (eqs. 16 and 17), the rate of cleavage of the carbon–hydrogen bonds is ca 10^{-10} times slower than the reaction of propagating radical, $\text{R}\cdot$, with oxygen.

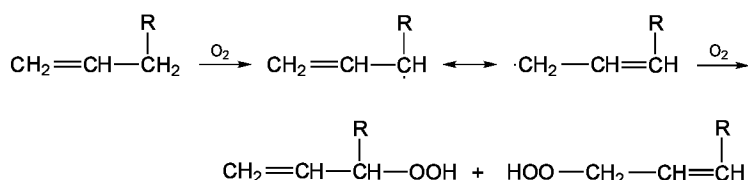


Therefore, steric factors and carbon–hydrogen bond energies significantly affect the autoxidation rate. The point of autoxidative hydrogen abstraction generally is that which forms the most stable alkyl radical. For example, a hydrogen on a tertiary carbon atom is abstracted more readily than a hydrogen on a primary carbon atom since the carbon–hydrogen BDEs are 381 and 406 kJ/mol (91 and 97 kcal/mol), respectively. The type and extent of termination reactions are influenced by substrate structure, oxygen pressure, and the presence of foreign substances, especially those with antioxidant properties.

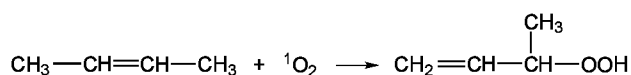
Autoxidations are usually carried out neat (with substrate as solvent), in a nonoxidizable solvent, or in an aqueous emulsion of the substrate. Because very high conversions generally are not obtained or desired, the hydroperoxide is removed from the oxidation stream by extraction with aqueous alkali.

Other compounds, eg, azoalkanes, acetone, etc, that yield alkyl radicals either thermally or by uv irradiation have been used with molecular oxygen to prepare alkyl hydroperoxides (r56).

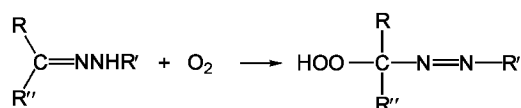
In the autoxidation of alkenes, different hydroperoxides are formed, depending on whether ground-state triplet oxygen or singlet oxygen is the oxidizing species. In the triplet state, oxygen abstracts an allylic hydrogen atom and the resulting allylic radical reacts with oxygen to form an allylic peroxyradical, which in turn propagates the oxidation chain by abstracting an allylic hydrogen atom from another olefin molecule (45) (eq. 18).



Singlet oxygen reacts with olefins presumably by the "ene" reaction to form allylic hydroperoxides (45,57), eg, 1-methyl-2-propenyl hydroperoxide [20733-08-8] is produced from 2-butene (eq. 19). The regioselectivity of this reaction has been investigated (58).



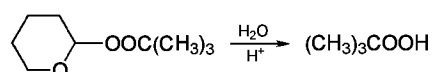
Hydroperoxides have been obtained from the autoxidation of alkanes, aralkanes, alkenes, ketones, enols, hydrazones, aromatic amines, amides, ethers, acetals, alcohols, and organomineral compounds, eg, Grignard reagents (10,45). In autoxidations involving hydrazones, double-bond migration occurs with the formation of hydroperoxy–azo compounds via free-radical chain processes (10,59) (eq. 20).



Commercially, autoxidation is used in the production of α -cumyl hydroperoxide, *tert*-butyl hydroperoxide, *p*-diisopropylbenzene monohydroperoxide, *p*-diisopropylbenzene dihydroperoxide, *p*-menthane hydroperoxide, pinane hydroperoxide, and ethylbenzene hydroperoxide.

Many organic peroxides of metals have been hydrolyzed to alkyl hydroperoxides. The alkylperoxy derivatives of aluminum, antimony, arsenic, boron, cadmium, germanium, lead, magnesium, phosphorus, silicon, tin, and zinc yield alkyl hydroperoxides upon hydrolysis (10,33,60,61).

Saponification of *tert*-alkyl peroxyesters yields alkyl hydroperoxides and carboxylic acids or their alkali metal salts. α -Ether-substituted peroxides can be hydrolyzed to the unsubstituted alkyl hydroperoxides, eg, *tert*-butyl hydroperoxide from *tert*-butyl 2-oxacyclohexyl peroxide [28627-46-5] (62):

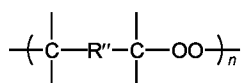


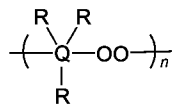
Other Hydroperoxides. Several hydrotroxides including alkyl hydrotroxides, $\text{R}-\text{OOOH}$, have been reported (63,64). There is strong spectroscopic evidence that α -cumyl hydrotroxide [82951-48-2] is produced in the low temperature ozonization of cumene. Homolytic decomposition of α -cumyl hydrotroxide in cumene–acetone-*d* in the presence of a hindered phenol resulted in cumyl alcohol as the only organic product (65). Based on the activation parameters obtained, α -cumyl hydrotroxide has a 10-h HLT of -26°C and a one-minute half-life temperature of 12°C , therefore α -cumyl hydrotroxide and other alkyl hydrotroxides, unlike alkyl hydroperoxides, are very unstable peroxides.

Little is known about the existence of alkyl hydrotetraoxides, $\text{R}-\text{OOOOH}$. There is some kinetic evidence supporting methyl hydrotetraoxide [23594-84-5] as a very labile intermediate in the reaction of methylperoxy radical, $\cdot\text{OO}\cdot$, and hydroperoxy radical, $\cdot\text{OOH}$ (63).

Dialkyl Peroxides

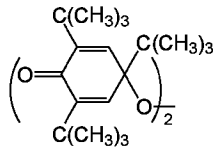
Dialkyl peroxides have the structural formula $\text{R}-\text{OO}-\text{R}'$, where R and R' are the same or different primary, secondary, or tertiary alkyl, cycloalkyl, and aralkyl hydrocarbon or hetero-substituted hydrocarbon radicals. Organomineral peroxides have the formulas $\text{R}_m\text{Q}(\text{OOR})_n$ and R_mQOOQR_m , where at least one of the peroxygens is bonded directly to the organo-substituted metal or metalloid, Q. Dialkyl peroxides include cyclic and bicyclic peroxides where the R and R' groups are linked, eg, endoperoxides and derivatives of 1,2-dioxane. Also included are polymeric peroxides, which usually are called poly(alkylene peroxides) or alkylene–oxygen copolymers, and poly(organomineral peroxides) (44), where Q = As or Sb.





Symmetrical dialkyl peroxides are commonly named as such, eg, dimethyl peroxide. For unsymmetrical dialkyl peroxides, the two radicals usually are listed in alphabetical order, eg, ethyl methyl peroxide. For organomineral peroxides or complex structures, ie, where R and R' are difficult to name as radicals, the peroxide is named as an alkylodioxy derivative, although alkylperoxy is still used by many authors. Cyclic peroxides are normally named as heterocyclic compounds, eg, 1,2-dioxane, or by substitutive oxa nomenclature, eg, 1,2-dioxacyclohexane; however, when the two oxygens form a bridge between two carbon atoms of a ring, the terms epidioxy or epiperoxy are frequently used. The resulting polycyclic structure has been called an endoperoxide, epiperoxide, or transannular peroxide.

Physical Properties. The structures and the boiling and melting points of several dialkyl peroxides are listed in Table 2; a comprehensive list is given in the literature (66). The melting point of 4,4'-dioxybis[2,4,6-tris(*tert*-butyl)-2,5-cyclohexadien-1-one] [1975-14-0] is 148–149°C.



Infrared, uv, nmr spectra (66), and photoelectron spectra have been reviewed (67). Physical properties of silicon peroxides are summarized in Reference 43. Other physical properties, eg, dipole moments, dihedral angles, and heats of combustion are listed in Reference 68. The oxygen–oxygen bond strengths of various dialkyl peroxides have been reported (69).

Table 2. Properties of Some Dialkyl Peroxides^a

Dialkyl peroxide	CAS Registry Number	Structure	Bp, °C (kPa) ^b	Mp, °C
dimethyl peroxide	[690-02-8]	CH ₃ –OO–CH ₃	13.5 (98.66)	
perfluoro dimethyl peroxide	[927-84-4]	CF ₃ –OO–CF ₃	37 (101.32)	
diethyl peroxide	[628-37-5]	C ₂ H ₅ –OO–C ₂ H ₅	62–63 (101.32)	
1,2-dioxane	[5703-46-8]		61.5 (14.67)	
<i>tert</i> -butyl methyl peroxide	[51392-67-7]	<i>t</i> -C ₄ H ₉ –OO–CH ₃	23 (2.53)	
<i>tert</i> -butyl 2-hydroxyethyl peroxide	[15476-85-4]	<i>t</i> -C ₄ H ₉ –OO–CH ₂ CH ₂ OH	37–38 (0.27)	
diisopropyl peroxide	[16642-57-2]	<i>i</i> -C ₃ H ₇ –OO– <i>i</i> -C ₃ H ₇		
3,3,5,5-tetramethyl-1,2-dioxolane	[22431-90-9]		55–58 (29.73), 46 (3.33)	14
di- <i>tert</i> -butyl peroxide	[110-05-4]	<i>t</i> -C ₄ H ₉ –OO– <i>t</i> -C ₄ H ₉	109 (101.32)	–18
perfluoro-di- <i>tert</i> -butyl peroxide	[26842-85-3]	(CF ₃) ₃ C–OO–C(CF ₃) ₃	99 (101.32)	
3,3,6,6-tetramethyl-1,2-dioxane	[22431-89-6]		44–45 (1.5) ^c	–26 ^c
di- <i>tert</i> -amyl peroxide	[10508-09-5]	<i>t</i> -C ₅ H ₁₁ –OO– <i>t</i> -C ₅ H ₁₁	44 (1.33)	
<i>tert</i> -butyl <i>tert</i> -cumyl peroxide	[3457-61-2]	<i>t</i> -C ₄ H ₉ –OO–C(CH ₃) ₂ C H ₅	40 (0.027)	13
9,10-dihydro-9,10-epidioxyanthracene	[4741-24-6]			120 ^d
2,5-dimethyl-2,5-di(<i>tert</i> -butylperoxy) hexane	[78-63-7]	[<i>t</i> -C ₄ H ₉ —OO— C(CH ₃) ₂ CH ₂ —] ₂	42 (0.008)	8
2,5-dimethyl-2,5-di(<i>tert</i> -butylperoxy)-3-hexyne	[1068-27-5]		65–67 (0.27)	
dicumyl peroxide	[80-43-3]	C H ₅ (CH ₃) ₂ C–OO–C(CH ₃) ₂ C H ₅		40–41
1,4-di(2- <i>tert</i> -butylperoxyisopropyl) benzene	[2781-00-1]	1,4-[<i>t</i> -C ₄ H ₉ —OO— C(CH ₃) ₂ —] ₂ C ₆ H ₄		79

^a Refs. 44 and 66.

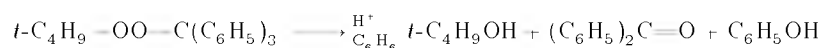
^b To convert kPa to mm Hg, multiply by 7.5.

^c Ref. 68.

^d Explodes at 120°C.

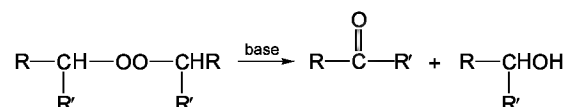
Di-*tert*-alkyl peroxides are thermally stable organic peroxides, having 10-h HLTs of 110–135°C for acyclic peroxides and 10-h HLTs approaching 200°C for five- to six-membered cyclic peroxides, eg, 129°C for di-*tert*-butyl peroxide (in decane) (22), 200°C for 3,3,5,5-tetramethyl-1,2-dioxolane [22431-90-9] (in benzene) (70), and 197°C for 3,3,6,6-tetramethyl-1,2-dioxane (in carbon tetrachloride) (71).

Metalloid peroxides behave as covalent organic compounds and most are insensitive to friction and impact but can decompose violently if heated rapidly. Most solid metalloid peroxides have well-defined melting points and the more stable liquid members can be distilled (Table 3). Some

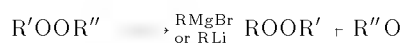


In Lewis acids, carbon–oxygen fission and Criegee-type rearrangements occur. For example, di-*tert*-butyl peroxide undergoes carbon–oxygen fission with BF_3 and Criegee-type rearrangement with AlCl_3 (66).

In the presence of base, di-*tert*-alkyl peroxides are stable, however primary and secondary dialkyl peroxides undergo oxygen–oxygen bond cleavage, forming alcohols, aldehydes, and ketones (44,66):



Dialkyl peroxides also undergo nucleophilic displacements by organometallic compounds:

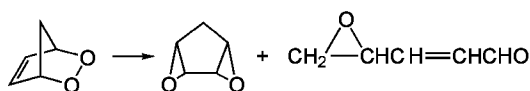


Primary and secondary dialkyl peroxides react much more readily than di-*tert*-alkyl peroxides (66,76). Products derived from the free radical are also produced in these reactions.

Substitution reactions on dialkyl peroxides without concurrent peroxide cleavage have been reported, eg, the nitration of dicumyl peroxide (44), and the chlorination of di-*tert*-butyl peroxide (77). Bromination by nucleophilic displacement on α -chloro- or α -hydroxyalkyl peroxides with hydrogen bromide produces α -bromoalkyl peroxides (78).

The polymeric peroxides, $\text{—(OOCH}_2\text{CXH—)}_n$, where X = H, C₆H₅, CH=CH₂, etc, are viscous liquids or amorphous solids having as many as 10 repeating units. These compounds usually explode when heated. The products obtained from the thermal or photodecomposition show that cleavage of both oxygen–oxygen and carbon–carbon bonds occurs. The type and amounts of products formed depend on the decomposition conditions and the structure of the peroxide. When X = phenyl, benzaldehyde and formaldehyde are the principal thermal decomposition products (79). Acids and bases also decompose polymeric peroxides. Oxygen–oxygen bond cleavage occurs under basic conditions, but the initially formed carbonyl compounds also are base-sensitive and are converted to other products. Reduction of polymeric peroxides, which proceeds more readily than that of dialkyl peroxides, has been carried out with a variety of reducing agents, eg, lithium aluminum hydride (44).

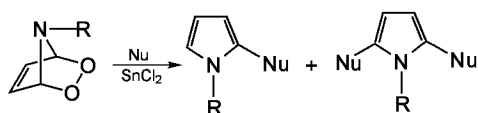
Unsaturated aliphatic endoperoxides form bis(epoxides) and α -epoxy aldehydes upon thermolysis (80,81). Thus 3,5-epidioxycyclopentene [6573-26-8] reacts as follows.



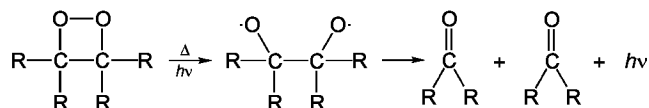
With certain transition metals, eg, Ru(II)-tertiary phosphine complexes, the principal products are bis(epoxides) (82).

The endoperoxides of polynuclear aromatic compounds are crystalline solids that extrude singlet oxygen when heated, thus forming the parent aromatic hydrocarbon (44,66,80,81). Thus 9,10-diphenyl-9,10-epidioxanthracene [15257-17-7] yields singlet oxygen and 9,10-diphenylanthracene.

Endoperoxides undergo carbon–oxygen cleavage in acids and oxygen–oxygen bond cleavage in bases, and they are more easily reduced than dialkyl peroxides. Reduction products depend on the structure and reducing agent. Aliphatic endoperoxides have been reduced, with and without involvement of an adjacent double bond that is present in most aliphatic endoperoxides, to glycols, epoxides, and hydrocarbons (44). Endoperoxides undergo hydrolysis or methanolysis to hydroperoxides (81). In the presence of stannous chloride, endoperoxides of N-substituted 1,2-dihydropyridines and N-substituted pyrroles react with carbon-centered nucleophiles (Nu) to give substituted products (81,83,84):



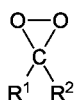
1,2-Dioxetanes have very low activation enthalpies (ca 109 kJ mol⁻¹), therefore, they are unstable at low temperatures and generally cleave thermally or photochemically at the oxygen–oxygen and carbon–carbon bonds. There is evidence that this fragmentation occurs through a labile 1,4-dioxy–diradical intermediate. Upon further fragmentation, chemiluminescence occurs and two carbonyl compounds are produced in the absence of trapping agents (66,80,85) (see LUMINESCENT MATERIALS, CHEMILUMINESCENCE). In equation 21, the R groups can be identical or dissimilar.



Based on reported activation parameters, tetramethyl-1,2-dioxetane [35856-82-7], where R = CH₃, has a 10-h HLT of about 15°C (85).

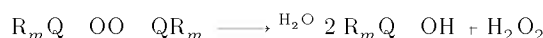
1,2-Dioxetanes are reduced to diols, epoxides, or allylic alcohols; the dioxetane structure and the reducing system determine which product forms or predominates (80). Although alkyl and aryl dioxetanes generally are inert to nucleophilic attack at the carbons, reactions with azides and caustic soda and rearrangements with Lewis acids have been reported (80). A variety of reactions involving 1,2-dioxetanes and heteroatoms have been observed, eg, rearrangements to α -hydroxyketones, nucleophilic additions, and rearrangements of 3,4-unsaturated-1,2-dioxetanes, ie, 1,2-dioxetenes (80). Reaction of 1,2-dioxetanes with diazoalkanes produces 1,3-dioxolanes and carbonyl compounds (86).

Dioxiranes are three-membered cyclic ring peroxides that are expected to



be very unstable owing to ring strain. Dimethyldioxirane [74087-85-7] where R¹ and R² = CH₃ has a 10-h HLT of about 44°C in acetone (87). Dioxiranes are effective oxygenating agents for epoxidations of olefins, allenes, polycyclic aromatic hydrocarbons, enols, and α,β -unsaturated ketones; for insertions of oxygen into X–H bonds of alkanes, primary and secondary alcohols, aldehydes, and silanes; and for oxidations of sulfides (to sulfoxides and sulfones), imines (to nitrones), and primary amines (to nitro compounds) (88–91). In these reactions, the dioxirane transfers oxygen to the substrate and generates the ketone from which the dioxirane was derived. Because of the weak oxygen–oxygen bond in dioxiranes, the mechanisms of these oxygen-transfer reactions are expected to have some diradical character. The autodecomposition of dimethyldioxirane reportedly involves radical intermediates such as methyl radical (92).

Most organomineral peroxides are hydrolytically unstable and readily hydrolyze to alkyl hydroperoxides or hydrogen peroxide (33,34,44,60,61):



Consequently, most organomineral peroxides must be prepared and stored under anhydrous conditions. In addition, anhydrous hydrogen chloride converts alkyl-substituted organomineral peroxides to alkyl hydroperoxides (33).

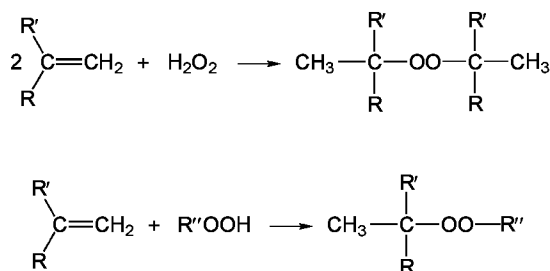
Basic hydrolysis of secondary alkyl-substituted silicon and germanium peroxides results in oxygen–oxygen bond cleavage.

The reduction of alkyl-substituted silicon and tin peroxides with sodium sulfite and triphenylphosphine has been reported (33,93). Alkyl-substituted aluminum, boron, cadmium, germanium, silicon, and tin peroxides undergo oxygen-to-metal rearrangements (33,43,94), eg, equations 22 and 23.



Organomineral peroxides also undergo thermal and photo-induced homolysis, yielding free radicals that are effective for initiating polymerization of vinyl monomers (44). Baeyer-Villiger oxidations of ketones to esters have been carried out using bis(trialkylsilyl)peroxides (95). Silane peroxides react with carbanions to generate products derived from attack either on silicon or on the oxygen–oxygen bond (96).

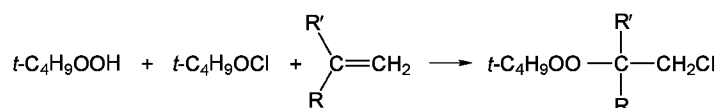
Synthesis. Dialkyl peroxides are prepared by the reaction of various substrates with hydrogen peroxide, hydroperoxides, or oxygen (69). They also have been obtained from reactions with other organic peroxides. For example, dialkyl peroxides have been prepared by the reaction of hydrogen peroxide and alkyl hydroperoxides with alkylating agents, eg, RX and olefins (33,66,97) (eqs. 24–27).



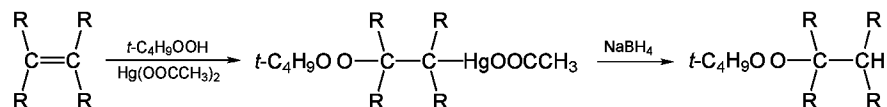
Reaction conditions depend on the reactants and usually involve acid or base catalysis. Examples of X include sulfate, acid sulfate, alkane- or arenesulfonate, chloride, bromide, hydroxyl, alkoxide, perchlorate, etc. RX can also be an alkyl orthoformate or *tert*-alkyl carboxylate. The reaction of cyclic alkylating agents, eg, epoxides and aziridines, with sodium or potassium salts of alkyl hydroperoxides also promotes formation of dialkyl peroxides (44,66). Olefinic alkylating agents include acyclic and cyclic olefinic hydrocarbons, vinyl and isopropenyl ethers, enamines, *N*-vinylamides, vinyl sulfonates, divinyl sulfone, and α , β -unsaturated compounds, eg, methyl acrylate, mesityl oxide, acrylamide, and acrylonitrile (44,66).

The following commercially available dialkyl peroxides are produced according to equations 24–27: di-*tert*-butyl peroxide from hydrogen peroxide and sulfated *tert*-butyl alcohol or isobutylene; dicumyl peroxide from α -cumyl hydroperoxide and cumyl alcohol, cumyl chloride, and α -methylstyrene; *m*- and *p*-di(2-*tert*-butylperoxyisopropyl)benzene [2781-00-2] from *tert*-butyl hydroperoxide [75-91-2] and *m*- and *p*-di(2-hydroxyisopropyl)benzene; 2,5-dimethyl-2,5-di(*tert*-butylperoxy)hexane; and 2,5-dimethyl-2,5-di(*tert*-butylperoxy)-3-hexyne from sulfated *tert*-butyl alcohol and 2,5-dimethyl-2,5-dihydroperoxyhexane and 2,5-dimethyl-2,5-dihydroperoxy-3-hexyne, respectively (97–99).

Olefins react with *tert*-butyl hydroperoxide in the presence of *tert*-butyl hypochlorite, forming *tert*-butyl β -chloroalkyl peroxides (66):



With mercuric acetate ($Hg(OOCCH_3)_2$), olefins and *tert*-butyl hydroperoxide form organomercury-containing peroxides (66,100). The organomercury compound can be treated with bromine or a mild reducing agent, such as sodium borohydride, to remove the mercury.



Primary and secondary alkyl halides and sulfonates react with potassium superoxide to form dialkyl peroxides (101,102) (eq. 28). Diazoalkanes, eg, diazomethane, have been used to alkylate hydroperoxides (66) (eq. 29).



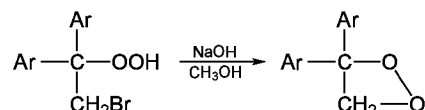
Unsymmetrical dialkyl peroxides are obtained by the reaction of alkyl hydroperoxides with a substrate, ie, R'H, from which a hydrogen can be abstracted readily in the presence of certain cobalt, copper, or manganese salts (eq. 30). However, this process is not efficient since two moles of the hydroperoxide are consumed per mole of dialkyl peroxide produced. In addition, side reactions involving free radicals produce undesired by-products (44,66).



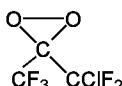
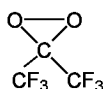
Symmetrical dialkyl peroxides have been prepared from alkyl hydroperoxides and lead tetraacetate. If tertiary dihydroperoxides are used, then cyclic

peroxides are obtained in high yields (44,66). Di-*tert*-butyl peroxide is obtained from the mild decomposition of di-*tert*-butyl diperoxyoxalate [1876-22-8] (44). Trifluoromethyl peroxide [927-84-4], CF_3OOCF_3 , is obtained from the photolysis of fluorofornyl peroxide and from the reaction of CF_3OF with COF_2 (66).

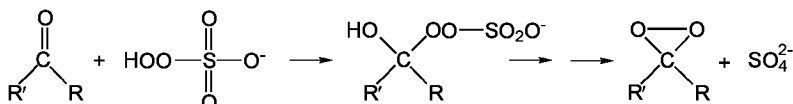
Olefins that are sufficiently electron-donating react with singlet oxygen by 1,2-addition to form 1,2-dioxetanes (80). Such electron-donating olefins are nitrogen heterocycles, *cis*- and *trans*-diethoxyethylene, tetramethoxyethylene, *p*-dioxene, 1,3-dioxole, and 2-methoxynorbornenes (66,85,93,103). 1,2-Dioxetanes are also obtained by the base-catalyzed rearrangement of 4-bromo-3-hydroxy-1,2-dioxolane derivatives and the ring closure of 1-bromo-2-hydroperoxyethane derivatives with methanolic caustic (Kopecky method) (85,104):



Two unstable and explosive dioxiranes, hexafluorodimethyldioxirane [35357-46-1] and chloropentafluorodimethyldioxirane [35357-48-3] have been synthesized (105).

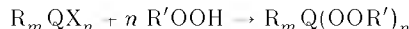


In the late 1970s, evidence showed that dialkyl dioxiranes were generated in ketone-caroate, $2\text{KHSO}_5 \cdot \text{KHSO}_4 \cdot \text{K}_2\text{SO}_4$, systems (106) and the mechanism of the reaction was determined (88,90):



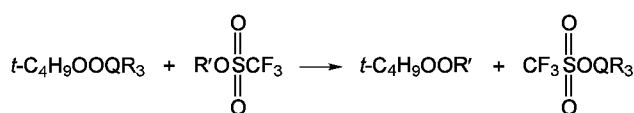
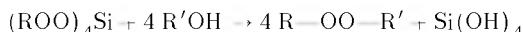
In the mid-1980s a convenient and relatively safe procedure was developed for vacuum codistillation of dimethyldioxirane and acetone from acetone-caroate systems (107). The resulting acetone solution of dimethyldioxirane could be used in subsequent oxygenation reactions (92).

Organomineral peroxides can be prepared by the reaction of certain organometallic or organometalloid compounds, R_mQX_n , with hydrogen peroxide or alkyl hydroperoxides:

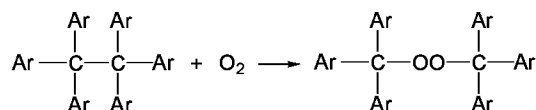


Organometallics and organometalloids that yield peroxides in this manner include those in which Q is aluminum, antimony, arsenic, boron, cadmium, germanium, lead, phosphorus, silicon, and tin and in which X is chlorine, bromine, alkoxy, acetoxy, cyano, oxide, hydride, hydroxyl, amino, alkyl, and boron tetrafluoride (28,33,44,60) (see Table 3).

Dialkyl peroxides may be prepared by reaction of alcohols or alkyl trifluoromethanesulfonates with organomineral peroxides of silicon, tin, and germanium (44,108), where Q = Sn and Ge:



Diaralkyl peroxides have been prepared by autoxidation. Those compounds which autoxidize to symmetrical diaralkyl peroxides form highly stabilized radical intermediates, eg, triphenylmethane, 9-phenylanthrone, and 2,4,6-tri(*t*-butyl)phenol (44,66). Compounds that form stable radicals by cleavage of carbon-carbon bonds can be autoxidized to diaralkyl peroxides (69).



Starting compounds include hexa- and pentarylethanes; the latter require higher temperatures (ca 100°C) than the former to break the carbon-carbon bond. In the presence of oxygen, stable radicals that are generated by other methods, eg, reduction of arylmethyl ethers and halides, also have been converted to diaralkyl peroxides (66).

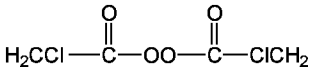
Autoxidation of alkanes generally promotes the formation of alkyl hydroperoxides, but di-*tert*-butyl peroxide has been obtained in >30% yield by the bromine-catalyzed oxidation of isobutane (66). In the presence of iodine, styrene also has been oxidized to the corresponding peroxide (44).

Olefins that polymerize readily in the presence of free radicals form peroxy polymers with oxygen: $\text{CH}_2=\text{CHX} + \text{O}_2 \rightarrow \text{-(OOCH}_2\text{CHX)-}_n$. Such copolymers of oxygen have been prepared from styrene, α -methylstyrene, indene, ketenes, butadiene, isoprene, 1,1-diphenylethylene, methyl methacrylate, methyl acrylate, acrylonitrile, and vinyl chloride (44,66,109). 1,3-Dienes, such as butadiene, yield randomly distributed 1,2- and 1,4-copolymers. Oxygen pressure and olefin structure are important factors in these reactions; for example, other products, eg, carbonyl compounds, epoxides, etc, can form at low oxygen pressures. Polymers possessing dialkyl peroxide moieties in the polymer backbone have also been prepared by base-catalyzed condensations of di(hydroxy-*tert*-alkyl) peroxides with dibasic acid chlorides or bis(chloroformates) (110).

Singlet oxygen adds 1,4- to 1,3-dienes, 1,3-endocyclic dienes, and polynuclear aromatics, forming cyclic peroxides (81). Compounds that yield these peroxides are rubrene, anthracene, substituted anthracenes, nitrogen heterocycles, ergosterol, α -terpinene, cyclopentadiene, 1,3-cycloheptadienes, *N*-phenyllophine, butadiene, and prostaglandin analogues (66,103,111-113).

Unsaturated transannular peroxides from cyclic dienes have been selectively reduced to the saturated peroxide analogues (114). For example,

1,3-cyclohexadiene is converted to 3,6-epidioxycyclohexene [6671-70-1] by singlet oxygen; then reduction leads to 1,4-epidioxycyclohexane [280-53-5]. Organomineral peroxides of antimony, arsenic, boron, magnesium, tin, cadmium, lead, silicon, and zinc have been prepared by autoxidation and some are listed in Table 3 (33,44,60,93,115). For example, dimethyl cadmium reacts with oxygen to form methylperoxy methyl cadmium [69331-62-0] and bis(methylperoxy) cadmium.

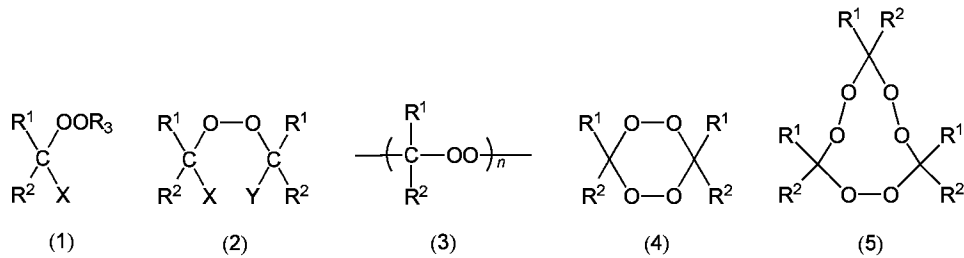


Thermally unstable cyclic trioxides, 1,2,3-trioxolanes or primary ozonides are prepared by reaction of olefins with ozone (64) (see OZONE). Dialkyl trioxides, ROOOR, have been obtained by coupling of alkoxy radicals, RO[•]; with alkylperoxy radicals, ROO[•]; at low temperatures. Dialkyl trioxides are unstable above -30°C (63). Dialkyl tetraoxides, ROOOOR, have been similarly produced by coupling of two alkylperoxy radicals, ROO[•]; at low temperatures. Dialkyl tetraoxides are unstable above -80°C (63).

α-Oxygen-Substituted Hydroperoxides and Dialkyl Peroxides

Dialkyl peroxides and hydroperoxides which have either a hydroxy, hydroperoxy, alkoxy, or alkylperoxy group on the carbon adjacent to the parent peroxide group are considered separately from the parent compounds due to their unique reactions and properties, but mainly because of their unique syntheses. Many of these compounds can be prepared by methods previously discussed for hydroperoxides and dialkyl peroxides, eg, autoxidation. However their primary preparation from aldehydes and ketones via reaction with hydrogen peroxide, alkyl hydroperoxides and peroxyacids is unique and makes it almost impossible to discuss them without referring to the parent carbonyl compound(s). These peroxides are often in equilibrium with other α-oxygen-substituted peroxides; they may be generated and used as chemical intermediates without being isolated (isolation may be either dangerous, impossible, or both). Unfortunately this means that the existence of some members of this group is speculative, based on products obtained from reactions of the peroxides rather than on isolation of the peroxides themselves.

The α-oxygen-substituted hydroperoxides and dialkyl peroxides comprise a great variety as shown in Figure 1. When discussing peroxides derived from ketones and hydrogen peroxide, (1) is often referred to as a ketone peroxide monomer and (2) as a ketone peroxide dimer.



several peroxide structures usually are present (4–6,10,44,117–121). Individual peroxides have been isolated from several ketones under different conditions (Table 5). The pure peroxides should be handled with extreme caution since most, especially those derived from the low molecular weight ketones, are shock- and friction-sensitive and can explode violently. Methyl ethyl ketone peroxide [1338-23-4] (MEKP) mixtures are produced commercially only as solutions containing <40 wt% MEKPs in solvents, commonly dialkyl phthalates.

Table 5. Melting Points of Some Peroxy Compounds from Ketones and Hydrogen Peroxide^a

Peroxy compound	CAS Registry Number	Structure	Mp, °C
2-chloro-1-hydroperoxy-cyclohexanol ^b	[15250-08-5]		76
1,1-dihydroperoxycyclo-dodecane ^c	[16623-96-4]		140
3,5-dihydroxy-3,5-dimethyl-1,2-dioxolane ^d	[37187-22-7]		90–91
di(1-hydroxycyclohexyl) peroxide ^e	[2407-94-5]		69–71
1-hydroxycyclohexyl 1-hydroperoxycyclohexyl peroxide	[78-18-2]	(6) $\begin{matrix} X & OHY & OOH \end{matrix}$	76–77
di(1-hydroperoxycyclohexyl) peroxide	[2699-12-9]	(6) $X - Y - OOH$	82–83
di(2-hydroperoxy-2-butyl) peroxide	[126-76-1]	(2) $R^1 - CH_3, R^2 - C_2H_5, X - Y - OOH$	39–42
3,3,6,6-tetramethyl-1,2,4,5-tetroxane	[1073-91-2]	(4) $R^1 - R^2 - CH_3$	131–133
3,6-diethyl-3,6-dimethyl-1,2,4,5-tetroxane ^f		(4) $R^1 - CH_3, R^2 - C_2H_5$	^f
7,8,15,16-tetraoxadispiro-[5.2.5.2]-hexadecane ^g	[183-84-6]		127–128
3,3,6,6,9,9-hexamethyl-1,2,4,5,7,8-hexoxononane	[17088-37-8]	(5) $R^1 = R^2 = CH_3$	96–97
3,6,9-triethyl-3,6,9-tri-methyl-1,2,4,5,7,8-hexoxononane	[24748-23-0]	(5) $R^1 - CH_3, R^2 - C_2H_5$	30–32
7,8,15,16,23,24-hexaoxatrispiro-[5.2.5.2.5.2] tetracosane ^h	[182-01-4]		93

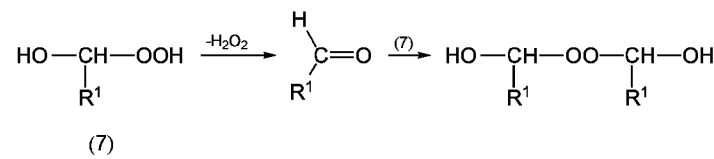
^a Refs. 10, 44, and 122.
^b Type (1) R¹ and R² are the ring; R³ = H; X = OH.
^c Type (1) R¹ and R² are the ring; R³ = H; X = OOH.
^d Type (2) R¹ = CH₃; R² is –CH₂–; X = Y = OH.
^e Structure (6) is type (2) wherein R¹ and R² are the ring and X and Y are specified.
^f The cis compound CAS Registry Number is [33817-91-3] with an mp of 12–14°C; the CAS Registry Number for the trans compound is [33817-92-4], mp 23–25°C.
^g Type (4) R¹ and R² are the ring.
^h Type (5) R¹ and R² are the ring.

Hydroxyalkyl Hydroperoxides. These compounds, represented by (1, X = OH, R³ = H), may be isolated as discreet compounds only with certain structural restrictions, eg, that one or both of R¹ and R² are hydrogen, ie, they are derived from aldehydes, or that R¹ or R² contain electron-withdrawing substituents, ie, they are derived from ketones bearing α-halogen substituents. Other hydroxyalkyl hydroperoxides may exist in equilibrium mixtures of ketone and hydrogen peroxide.

Hydroxyalkyl hydroperoxides which have been isolated are moderately stable and many are unaffected by distillation. However, the low molecular weight members, eg, hydroxymethyl hydroperoxide (1, X = OH, R¹ = R² = R³ = H), can decompose violently. 1,1,1,3,3,3-Hexafluoro-2-hydroxy-2-propyl hydroperoxide [32751-01-2] (1), where X = OH, R³ = H, and R¹ = R² = CF₃, is a viscous liquid that decomposes slowly at 25°C (123). The melting points of some hydroxyalkyl hydroperoxides are listed in Tables 4 and 5. Thermal decomposition yields mixtures of carbonyl compounds, carboxylic acids, and alcohols. These hydroxyhydroperoxides have been reduced to the corresponding carbonyl compounds or alcohols with a variety of reducing agents including hydriodic acid, phosphines, zinc and acetic acid, and by catalytic hydrogenation. Upon reaction with lead tetraacetate, hydroxyalkyl hydroperoxides, like other hydroperoxides, liberate oxygen.

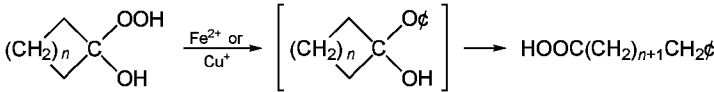
Hydroxyalkyl hydroperoxides from cyclic ketones (1), where X = OH, R³ = H and R¹, R² = alkylene, apparently exist in solution as equilibrium mixtures of the cyclic ketone, hydrogen peroxide, and other peroxides, eg, the dihydroperoxide (1) in which X = OOH, and dialkyl peroxides (2) where X = OH and Y = OH or OOH. Due to the existence of this equilibrium, the latter two dialkyl peroxides react as mixtures of monomeric hydroperoxides in solution.

Hydroxyalkyl hydroperoxides having at least one α-hydrogen ie, (7, X = OH, R = alkyl, R² = R³ = H), ie, those derived from aldehydes, lose hydrogen peroxide and form dialkyl peroxides (2, X = Y = OH), especially in the presence of water:



Acidic hydrolysis of these hydroxyalkyl hydroperoxides yields carboxylic acids, whereas basic hydrolysis regenerates the parent aldehyde, hydrogen peroxide, and often other products. When derived from either aldehydes or cyclic ketones, peroxides (**1**, X = OH, R³ = H, R¹, R² = alkylene or R¹ = alkyl, R² = H) can be condensed, using acid, dehydrating agents (eg, phosphorus pentoxide), or vacuum to form polymeric peroxides (**3**). Most such polymeric peroxides decompose explosively.

As with other hydroperoxides, hydroxyalkyl hydroperoxides are decomposed by transition-metal ions in an electron-transfer process. This is true even for those hydroxyalkyl hydroperoxides that only exist in equilibrium. For example, those hydroperoxides from cyclic ketones (R¹, R² = alkylene) form an oxygen-centered radical initially which then undergoes ring-opening β-scission forming an intermediate carboxyalkyl radical (124):



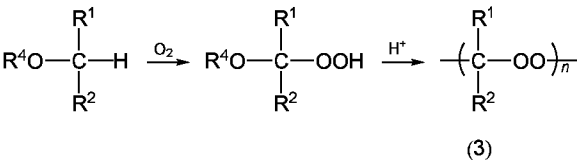
Without other alternatives, the carboxyalkyl radicals couple to form dibasic acids; HOOC(CH₂)²ⁿ⁺⁴COOH. In addition, the carboxyalkyl radical can be used for other desired radical reactions, eg, hydrogen abstraction, vinyl monomer polymerization, addition of carbon monoxide, etc. The reactions of this radical with chloride and cyanide ions are used to produce amino acids and lactams employed in the manufacture of polyamides, eg, nylon.

Secondary alcohols, such as isopropyl alcohol, *sec*-butyl alcohol, 2-pentanol, 3-pentanol, cyclopentanol, and cyclohexanol, have been autoxidized to hydroxyalkyl hydroperoxides (**1**, X = OH; R³ = H) (10,44). These autoxidations usually are carried out at ca 20°C with uv radiation in the presence of a photosensitizer, eg, benzophenone. α-Oxygen-substituted dialkyl peroxides (**2**, X = Y = OH and X = Y = OOH), also are formed and sometimes they are the exclusive products (10).

Alkoxyalkyl Hydroperoxides. These compounds (**1**, X = OR⁴, R³ = H) have been prepared by the ozonization of certain unsaturated compounds in alcohol solvents (10,125,126). 2-Methoxy-2-hydroperoxypropane [10027-74-4] (**1**, X = OR⁴, R⁴ = methyl), has been generated in methanol solution and spectral data obtained (127). A rapid exothermic decomposition upon concentration of this peroxide in a methylene chloride–methanol solution at 0°C has been reported (128). 2-Bromo-1-methoxy-1-methylethyl hydroperoxide [98821-14-8] has been distilled (bp 60°C (bath temp.), 0.013 kPa) (129). Two cyclic alkoxyalkyl hydroperoxides from cyclodecanone have been reported (**1**, where X = OR⁴; R¹, R² = 5-oxo-1,9-nonanediyl) with mp 94–95°C (R⁴ = methyl) and mp 66–68°C (R⁴ = ethyl) (130). Like other hydroperoxides, alkoxyalkyl hydroperoxides can be acylated or alkylated (130,131).

α-Trialkylsiloxyhydroperoxides (**1**, X = OSiR₃, R³ = H) have been prepared by the reaction of appropriate silyl enol ethers with hydrogen peroxide in the presence of acid (132).

Alkoxyalkyl hydroperoxides are more commonly called ether hydroperoxides. They form readily by the autoxidation of most ethers containing α-hydrogens, eg, dioxane, tetrahydrofuran, diethyl ether, diisopropyl ether, di-*n*-butyl ether, and diisooamyl ether (10,44). From certain ethers, eg, diethyl ether (in the following, R¹ = H; R² = CH₃; R⁴ = CH₂CH₃), the initially formed ether hydroperoxide can yield alcohol on standing, or with acid treatment form dangerously shock-sensitive and explosive polymeric peroxides (**3**).



Low molecular weight ether hydroperoxides are similarly dangerous and therefore ethers should be tested for peroxides and any peroxidic products removed from them before ethers are distilled or evaporated to dryness. Many ethers autoxidize so readily that peroxidic compounds form at dangerous levels when stored in containers that are not airtight (133). Used ether containers should be handled cautiously and if they are found to contain hazardous solid ether peroxides, bomb-squad assisted disposal may be required (134). Zeolites have been used for removal of peroxide impurities from ethers (135).

Hydroxyalkyl Alkyl Peroxides and Hydroxyalkyl Peroxyesters. The same structural restrictions discussed for the hydroxyhydroperoxides apply for the hydroxyalkyl alkyl peroxides, and those that exist are derived from aldehydes and certain ketones having electron-withdrawing groups, eg, polyfluorinated α,β-unsaturated ketones (136).

Hydroxyalkyl alkyl peroxides (**1**, X = OH, R³ = alkyl) are reasonably stable and usually can be distilled under a vacuum; the boiling points and structures of representative compounds are listed in Table 6. Low molecular weight compounds, eg, hydroxymethyl methyl peroxide, are unstable and explosive. Reactions with hydriodic acid or titanium trichloride generally are not quantitative but reductions to carbonyl- and hydroxy-containing compounds have been reported (44). Hydrolysis of hydroxyalkyl alkyl peroxides with α-hydrogen (ie, those derived from aldehydes) usually generate a carbonyl compound and the corresponding alkyl hydroperoxide. In alkali, hydrolysis of hydroxymethyl alkyl peroxides liberates hydrogen, formic acid, formaldehyde, and alcohol from R³. Thermal decompositions of hydroxyalkyl alkyl peroxides appear to involve both homolytic and heterolytic processes. Ferrous salts can be used to generate free radicals (4,5,44).

Table 6. Boiling Points of Some Hydroxyalkyl Alkyl Peroxides^a

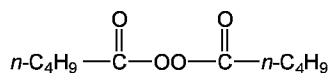
Hydroxyalkyl alkyl peroxide	CAS Registry Number	R ³	R ¹	Bp, °C (kPa) ^b
hydroxymethyl methyl peroxide	[40116-50-5]	CH ₃	H	45 (2.27)
<i>tert</i> -butyl hydroxy-methyl peroxide	[17742-78-8]	<i>t</i> -C ₄ H ₉	H	52–53 (1.07)
1-hydroxyethyl methyl peroxide	[28567-26-2]	CH ₃	CH ₃	25–27 (2.27)
1-hydroxyethyl ethyl peroxide	[28567-28-4]	C ₂ H ₅	CH ₃	48–50 (8.67)
<i>tert</i> -butyl 1-hydroxy-ethyl peroxide	[4202-06-6]	<i>t</i> -C ₄ H ₉	CH ₃	30–31.5 (0.13)
<i>tert</i> -butyl 1-hydroxy-butyl peroxide	[13258-56-5]	<i>t</i> -C ₄ H ₉	<i>n</i> -C ₃ H ₇	34–37 (0.13)

^a Figure 1, structure (**1**), R² = H; X = OH; R¹ and R³ are specified.

^b To convert kPa to mm Hg, multiply by 7.5.

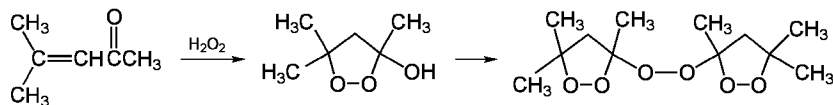
Hydroxyalkyl peroxyesters (**1**, X = OH, R¹, R² = hydrogen, alkyl; R³ = acyl) are proposed as intermediates in the Baeyer-Villiger oxidation of

aldehydes and ketones to esters by peroxy-carboxylic acids. Such a peroxyester has been isolated from the reaction of peroxyacetic acid and acetaldehyde (137):



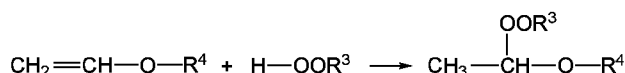
Hydroxyalkyl peroxyesters also have been isolated from the autoxidation products of aldehydes and by esterification of hydroxyhydroperoxides (44).

Mesityl oxide and hydrogen peroxide react initially to form the cyclic hydroxyalkyl alkyl peroxide, a 1,2-dioxolane. Prolonged equilibration results in formation of the cyclic di(alkylperoxyalkyl) peroxide, 3,3'-dioxibis(3,5,5-trimethyl-1,2-dioxolane) [4507-98-6] (122,138):



In the presence of strong acid catalysts such as sulfuric acid, aliphatic (R^1CHO) aldehydes react with alkyl hydroperoxides, eg, *tert*-alkyl hydroperoxides, to form hydroxyalkyl alkyl peroxides (1), where $\text{X} = \text{OH}$; R^1, R^2 = hydrogen, alkyl; and R^3 = *tert*-alkyl.

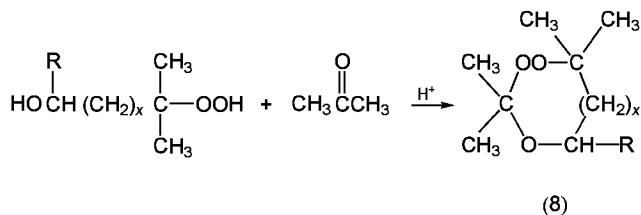
Alkoxyalkyl Alkyl Peroxides. *tert*-Butyl tetrahydropyran-2-yl peroxide [28627-46-5] (1), where R^3 = *tert*-butyl, $\text{X} = \text{OR}^4$, $\text{R}^1 = \text{H}$, R^2 and R^4 = 1,4-butanediyl, has been isolated. This is one of many examples of alkoxyalkyl alkyl peroxides which may be prepared by reaction of hydroperoxides with vinyl ethers (139):



2-Methoxy-1-methylethyl alkyl peroxides (1, $\text{X} = \text{OR}^4$, $\text{R}^1 = \text{R}^2 = \text{R}^4$ = methyl, R^3 = decyl, undecyl, pentadecyl, 2-octyl) are liquids that have been isolated using flash chromatography (131). Peroxyester derivatives (R^3 = acyl) have also been prepared (131).

1,2,4-Trioxacycloalkanes. 1,2,4-Trioxanes (1, $\text{X} = \text{OR}^4$; R^3 and R^4 = alkylene) are generally prepared by the interaction of aldehydes with zwitterionic intermediates made from reaction of singlet oxygen with olefins. They can also be prepared by catalyzed reaction of ketones or aldehydes with 1,2-dioxetanes or endoperoxides, and they can be prepared directly from certain hydroperoxides. Their preparation, molecular structure, and reactivity have been reviewed (140). 1,2,4-Trioxan-5-ones (R^3, R^4 = oxoethylene) are prepared by reaction of ketones or aldehydes with trimethylsilyl- α -trimethylsilylperoxycarboxylates (141).

1,2,4-Trioxacycloheptanes ($x = 1$) and 1,2,4-trioxacyclooctanes ($x = 2$) are synthesized by the reaction of suitable hydroxyhydroperoxides with aldehydes and ketones in the presence of acid catalysts (17). Cyclic peroxides are significantly more stable than the analogous acyclic peroxides; eg, 3,3,5,7,7-pentamethyl-1,2,4-trioxacycloheptane (8), where $x = 1$ and $\text{R} = \text{CH}_3$, is characterized by a 10-h HLT of 173°C in benzene (142).

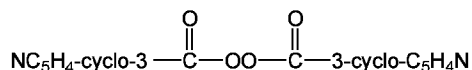


Geminal Dihydroperoxides. These dihydroperoxides (1, $\text{X} = \text{OOH}$, $\text{R}^3 = \text{H}$) can be made from many different carbonyl compounds. The structural restrictions discussed for hydroxyalkyl hydroperoxides generally do not apply. These peroxides can also be synthesized by perhydrolysis of ketals (143).

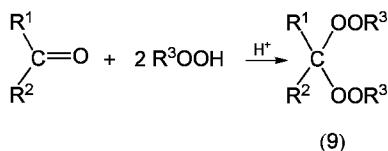
Low molecular weight dihydroperoxides (1, $\text{X} = \text{OOH}$; $\text{R}^3 = \text{H}$, R^1 and R^2 = short-chain alkyl) are soluble in water and are explosive when pure. They have been reduced to the corresponding ketones with hydriodic acid or zinc and acetic acid. Hydrolysis also gives the corresponding ketones. In the presence of catalytic amounts of acids or on prolonged storage, solutions of dihydroperoxides form equilibrium amounts of hydrogen peroxide and di(hydroperoxyalkyl) peroxides (2, $\text{X} = \text{Y} = \text{OOH}$) and ultimately equilibrium amounts of cyclic triperoxides (5). Oxygen evolution and formation of polymeric peroxides (3) also are observed from dihydroperoxides, especially in the presence of impurities, eg, trace transition metals, bases, etc. In the presence of sulfuric acid, dihydroperoxides can form cyclic diperoxides (4) via the first-formed cyclic triperoxide (5) (118). With cold formic acid, dihydroperoxides are converted to esters, apparently by a Baeyer-Villiger-like reaction (44). Dihydroperoxides have been characterized by esterification to solid diperoxyester derivatives (144) and by hydriodic acid titration.

Thermal decomposition of dihydroperoxides results in initial homolysis of an oxygen–oxygen bond followed by carbon–oxygen and carbon–carbon bond cleavages to yield mixtures of carbonyl compounds (ketones, aldehydes), esters, carboxylic acids, hydrocarbons, and hydrogen peroxide. Commercially available MEKP formulations are mixtures of the dihydroperoxide (1), where $\text{X} = \text{OOH}$; $\text{R}^3 = \text{H}$, $\text{R}^1 =$ methyl, and $\text{R}^2 =$ ethyl (2,2-dihydroperoxybutane [2625-67-4]), and dialkyl peroxide (2), where $\text{X} = \text{OOH}$, $\text{Y} = \text{OOH}$, $\text{R}^1 =$ methyl, and $\text{R}^2 =$ ethyl (di(2-hydroperoxy-2-butyl) peroxide [126-76-1]). These formulations are widely used as free-radical initiators in the metal-promoted cure of unsaturated polyester resins at about 20°C.

When derived from cyclic ketones, dihydroperoxides (1, $\text{X} = \text{OOH}$; $\text{R}^3 = \text{H}$, R^1 and R^2 = alkylene) react with vinyl monomers in the presence of Cr, V, or Ti salts to form difunctional compounds (145):



Diperoxyketals and Diperoxyacetals. Aromatic aldehydes react with alkyl hydroperoxides in the presence of strong acid catalysts such as sulfuric acid to form diperoxyacetals (1, $\text{X} = \text{OOR}^5$; $\text{R}^1 = \text{H}$, $\text{R}^2 = \text{Ar}$, $\text{R}^3 = \text{R}^5 =$ alkyl). Diperoxyketals (1, $\text{X} = \text{OOR}^5$; $\text{R}^1, \text{R}^2, \text{R}^3, \text{R}^5 =$ alkyl) are generally prepared by acid-catalyzed reaction of a ketone with two equivalents of an alkyl hydroperoxide. There are few structural limitations on the R^1 and R^2 groups. Aromatic ketones react with alkyl hydroperoxides only under extremely strong acid-dehydrating conditions to generate the diperoxyketal (9, $\text{R}^1 = \text{Ar}$, $\text{R}^2 = \text{Ar}$ or alkyl).



Diperoxyketals are solids or colorless liquids and are soluble in common organic solvents and insoluble in water. The physical properties and structures of some diperoxyketals are listed in Table 7. In the pure state, the low molecular weight compounds can decompose violently when heated, and addition of concentrated sulfuric acid can result in flaming decompositions. There are many commercial diperoxyketals, and they are usually diluted with solvents for improved safety. Diperoxyketals can be analyzed using hydriodic acid or stannous chloride even in the presence of less reactive dialkyl peroxides. Diperoxyketals have been reduced to alcohols with hydrogen and Raney nickel (44). Acid hydrolysis yields a ketone and alkyl hydroperoxide. Diperoxyketals undergo exchange reactions with hydroperoxides; such reactions are used to synthesize cumylperoxy derivatives (44):

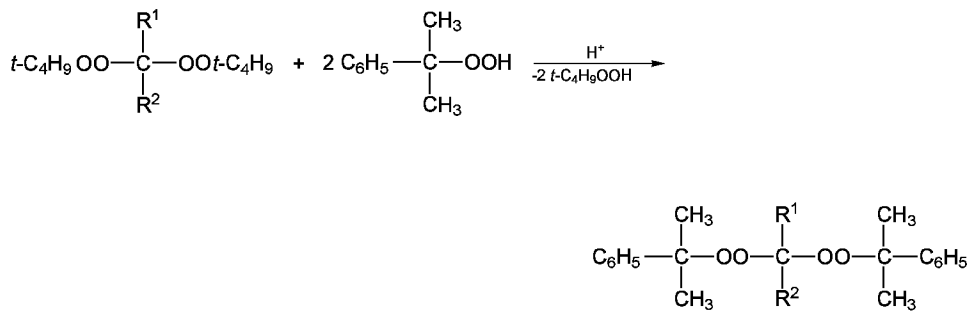
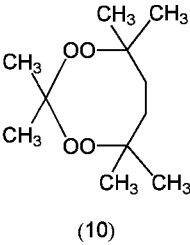


Table 7. Boiling Points of Some Diperoxyketals^a

Diperoxyketal	CAS Registry Number		Bp, °C (kPa) ^b
2,2-di(<i>t</i> -butyl-peroxy)propane	[4262-61-7]		69–70 (2.0)
2,2-di(<i>t</i> -amyl-peroxy)propane	[3052-70-8]		68 (0.17)
2,2-di(<i>t</i> -butyl-peroxy)butane	[2167-23-9]		50 (0.27)
1,1-di(<i>t</i> -butyl-peroxy)cyclohexane	[3006-86-8]		52–54 (0.02)
2,2-bis[4,4-di(<i>t</i> -butyl-peroxy)cyclo-hexyl]-propane	[1705-60-8]		117–120 ^c

^a Refs. 4 and 44.
^b To convert kPa to mm Hg, multiply by 7.5.
^c Mp, °C; Ref. 72.

In the presence of strong acid catalysts many commonly used commercial *tert*-alkyl hydroperoxides decompose to acetone to some extent. Consequently, the diperoxyketals derived from other ketones and *tert*-alkyl hydroperoxides are often contaminated with small amounts of diperoxyketals derived from acetone (**1**, X = OOR⁵, R¹ = R² = methyl, R³ = R⁵ = *tert* alkyl). Tertiary diperoxyketals (**1**, X = OOR⁵, R¹, R² = alkyl, R³, R⁵ = tertiary alkyl) are excellent free-radical initiators. Such diperoxyketals are stable, and those with R³ = R⁵ = *tert* butyl have 10-h HLTs in the 93–114°C range. Less thermally stable diperoxyketals are those derived from cyclic ketones and those with bulkier *tert*-alkyl groups, eg, *tert*-amyl, *tert*-octyl, *tert*-cumyl. Commercial members of this group all have R³ = R⁵, and thermally decompose to free radicals by cleavage of only one oxygen–oxygen bond initially, usually followed by β-scission of the resulting alkoxy radicals (146). For acyclic diperoxyketals, β-scission produces an alkyl radical and a peroxyester. Owing to similarity of thermal stability, the peroxyester decomposes almost simultaneously. With cyclic diperoxyketals, such as 1,1-di(*tert*-butylperoxy)cyclohexane, β-scission cleaves the cycloalkyl ring to give an alkyl radical with an attached peroxyester group. The effect, after peroxyester decomposition, is the production of two monoradicals, a diradical, and carbon dioxide (147). In the presence of polymerizable monomers, certain cyclic diperoxyketals partially decompose and the peroxyester moiety becomes a polymeric end group, thus forming a peroxypolymer (148). Diperoxyketals, and many other organic peroxides, are acid-sensitive, therefore removal of all traces of the acid catalysts must be accomplished before attempting distillations or kinetic decomposition studies. The low molecular weight diperoxyketals can decompose with explosive force and commercial formulations are available only as mineral spirits or phthalate ester solutions. Diperoxyketals in which the peroxide groups are in a ring are among the most thermally stable organic peroxides, eg, 3,3,6,6,9,9-hexamethyl-1,2,4,5-tetraoxacydononane [22397-33-7] (**10**) which has 10-h HLT = 141°C in benzene.

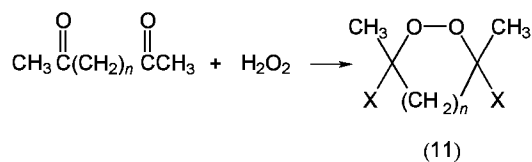


Di(hydroxyalkyl) Peroxides. The lowest molecular weight member of this group (**2**, X = Y = OH), di(hydroxymethyl) peroxide (R¹ = R² = OH) is a dangerously explosive solid. With increasing molecular weight, di(hydroxyalkyl) peroxides become liquids and eventually solids of

decreasing explosive nature and water solubility. In solution, these dialkyl peroxides exist in equilibrium with other α -oxygen-substituted peroxides, carbonyl compounds, and hydrogen peroxide (44). The existence of several equilibrium components explains some chemical properties of these mixtures, eg, formation of dibasic acids with ferrous salts or by thermal decomposition of cyclic peroxides, and the quantitative liberation of iodine from acidified iodides. Di(hydroxyalkyl) peroxides have been reduced to carbonyl- and hydroxy-containing derivatives with zinc and acetic acid, lithium aluminum hydride, sodium, sulfur dioxide, and ferrous salts. Hydrolysis can occur in neutral, acidic, or basic solutions and yields the corresponding carbonyl compounds, hydrogen peroxide, carboxylic acids, and hydrogen (44).

Formaldehyde reacts with di(hydroxymethyl) peroxide and phosphorus pentoxide to form di(hydroxymethoxymethyl) peroxide (2), where $X = Y = OCH_2OH$, $R^1 = R^2 = H$ (122).

Reaction of 1,3- and 1,4-diketones ($n = 1$ or 2) with hydrogen peroxide yields cyclic di(hydroxyalkyl) ($X = OH$) or di(hydroperoxyalkyl) ($X = OOH$) peroxides (11) (122).



The di(hydroxyalkyl) peroxide (2) from cyclohexanone is a solid which is produced commercially. The di(hydroxyalkyl) peroxide (2) from 2,4-pentanedione (11, $n = 1$; $X = OH$) is a water-soluble solid which is also produced commercially (see Table 5). Both these peroxides are used for curing cobalt-promoted unsaturated polyester resins. Because these peroxides are susceptible to promoted decomposition with cobalt, they must exist in solution as equilibrium mixtures with hydroperoxide structures (122,149).

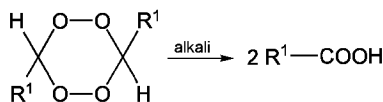
Hydroxyalkyl Hydroperoxyalkyl Peroxides. There is evidence that hydroxyalkyl hydroperoxyalkyl peroxides (2, $X = OH$, $Y = OOH$) exist in equilibrium with their corresponding carbonyl compounds and other α -oxygen-substituted peroxides. For example, reaction with acyl halides yields diperoxyesters. Dilute acid hydrolysis yields the corresponding ketone (44). Reduction with phosphines yields di(hydroxyalkyl) peroxides and dehydration results in formation of cyclic diperoxides (4).

Thermal decomposition of hydroxyalkyl hydroperoxyalkyl peroxides produces mixtures of starting carbonyl compounds, mono- and dicarboxylic acids, cyclic diperoxides, carbon dioxide, and water. One specific hydroxyalkyl hydroperoxyalkyl peroxide from cyclohexanone (2, $X = OH$, $Y = OOH$) is a solid that is produced commercially as a free-radical initiator and bleaching agent (see Table 5). On controlled decomposition, it forms 1,12-dodecanedioic acid (150).

Di(hydroperoxyalkyl) Peroxides. Low molecular weight di(hydroperoxyalkyl) peroxides (2, $X = Y = OOH$) are dangerously prone to explosive decomposition when they are pure. Some have been characterized by acylation to the corresponding diperoxyesters (2, $X = Y = OOR^5$, $R^5 = \text{acyl}$), eg, with *p*-nitrobenzoyl chloride. Upon reaction with lead tetraacetate, di(hydroperoxyalkyl) peroxides can also be converted to cyclic diperoxides (4). They are also converted to symmetrical or unsymmetrical cyclic triperoxides (5) in the presence of a second ketone and a catalyst, eg, $\text{CuSO}_4\text{-HCl}$ (44,119).

Cyclic Peroxides. Cyclic diperoxides (4) and triperoxides (5) are solids and the low molecular weight compounds are shock-sensitive and explosive (151). The melting points of some characteristic compounds of this type are given in Table 5. They can be reduced to carbonyl compounds and alcohols with zinc and alkali, zinc and acetic acid, aluminum amalgam, Grignard reagents, and warm acidified iodides (44,122). They are more difficult to analyze by titration with acidified iodides than the acyclic peroxides and have been successfully analyzed by gas chromatography (112).

Acid hydrolysis of peroxides (4) and (5) generates carbonyl compounds (parent ketones or aldehydes) and hydrogen peroxide. Basic hydrolysis of cyclic diperoxides with α -hydrogen (those from aldehydes) yields carboxylic acids (44):



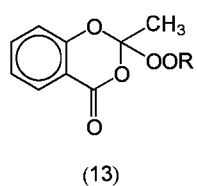
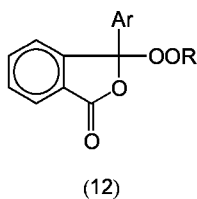
Thermal and photochemical decomposition of peroxides (4) and (5) lacking α -hydrogens (those derived from ketones) produces macrocyclic hydrocarbons and lactones (119,152,153). For example, 7,8,15,16,23,24-hexaoxatrispiro [5.2.5.2.5.2] tetracosane (see Table 5) yields cyclopentadecane and oxacycloheptadecan-2-one.

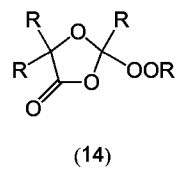
Stereoisomers of peroxides (4) and (5) are known to exist and their conformations have been studied using photoelectron, nmr, and other spectroscopic methods, and their crystalline structures have been determined (122,154,155).

Polymeric α -Oxygen-Substituted Peroxides. Polymeric peroxides (3) are formed from the following reactions: ketone and aldehydes with hydrogen peroxide, ozonization of unsaturated compounds, and dehydration of α -hydroxyalkyl hydroperoxides; consequently, a variety of polymeric peroxides of this type exist. Polymeric peroxides are generally viscous liquids or amorphous solids, are difficult to characterize, and are prone to explosive decomposition.

Polymeric *gem*-peroxides (3) from hydrogen peroxide and lower carbon ketones have been separated by paper or column chromatography and have been characterized by conversion to the bis(*p*-nitro)peroxybenzoates. Oligomeric peroxides (3, $R^1 = \text{methyl}$, $R^2 = \text{ethyl}$, $n = 1-4$) from methyl ethyl ketone have been separated and interconverted by suitable treatment with ketone and hydrogen peroxide (44).

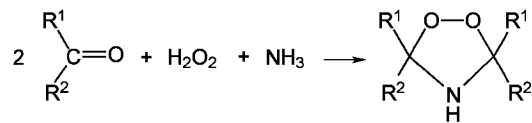
Miscellaneous α -Substituted Peroxides. 3-Aryl-3-(*tert*-alkylperoxy)-phthalides (12) are prepared from the corresponding 3-chlorophthalides and *tert*-alkyl hydroperoxide (156). 2-Methyl-2-(*tert*-alkylperoxy)-1,3-benzodioxan-4-ones (13) are obtained from *o*-acetylsalicyloyl chloride and *tert*-alkyl hydroperoxides (157). Trisubstituted 2-(*tert*-alkylperoxy)-1,3-dioxolan-4-ones (14) are synthesized from sterically favored α -acyloxy acid chlorides and *tert*-alkyl hydroperoxides (158).



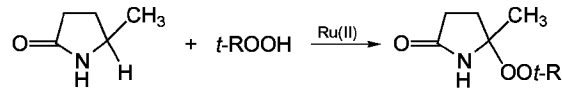


1-(*tert*-Butyldimethylsiloxy)-1-methoxyhexyl hydroperoxide [87258-13-7] has been synthesized from the corresponding ester-derived silyl ketene acetal and anhydrous hydrogen peroxide (132).

Certain ketones, eg, cyclohexanone, react with hydrogen peroxide in the presence of ammonia, yielding 1,2,4-dioxazolidines (159):

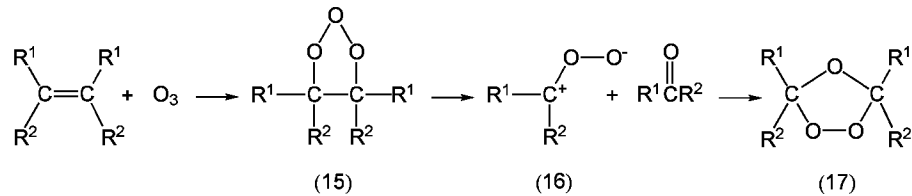


α -Amido-dialkyl peroxides have been prepared by ruthenium-catalyzed oxidation of amides with *t*-alkyl hydroperoxides (160):

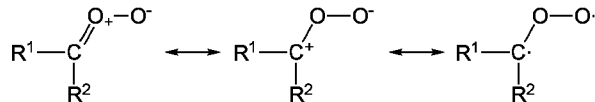


Ozonides and Ozonization

Unsaturated compounds undergo ozonization to initially produce highly unstable primary ozonides (15), ie, 1,2,3-trioxolanes, also known as molozonides, which rapidly split into carbonyl compounds (aldehydes and ketones) and 1,3-zwitterion (16) intermediates. The carbonyl compound-zwitterion pair then recombines to produce a thermally stable secondary ozonide (17), also known as a 1,2,4-trioxolane (44,64,125,161,162).

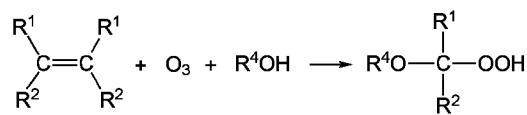


The 1,3-zwitterion appears to have some diradical as well as 1,2-zwitterionic (carbonyl oxide) character:



Most ozonolysis reaction products are postulated to form by the reaction of the 1,3-zwitterion with the extruded carbonyl compound in a 1,3-dipolar cycloaddition reaction to produce stable 1,2,4-trioxanes (ozonides) (17) as shown; with itself (dimerization) to form cyclic diperoxides (4); or with protic solvents, such as alcohols, carboxylic acids, etc, to form α -substituted alkyl hydroperoxides. The latter can form other peroxidic products, depending on reactants, reaction conditions, and solvent.

In the presence of alcohols, the ozonization products are alkoxyalkyl hydroperoxides (1, X = OR⁴, R³ = H):



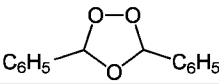
By-products include ozonides (17). Other peroxidic products including polymeric peroxides and polymeric ozonides can form, depending on reaction conditions, solvent, and olefin used. A variety of cyclic diperoxides (4) have been obtained by ozonolysis of olefins. Both *cis*- and *trans*-1,2,4-trioxanes (17) are formed when asymmetrical internal olefins are ozonized. Boiling point data for several 1,2,4-trioxanes are listed in Table 8. Ozonolysis of asymmetrical olefins produces asymmetrical and symmetrical 1,2,4-trioxanes, ie, cross ozonides form. This is due to generation of two different 1,3-zwitterion intermediates and two different carbonyl compounds from the asymmetrical primary ozonide. 1,2,4-Trioxanes also have been synthesized by dehydration of di(hydroxyalkyl) peroxides (2), where X = Y = OH, with phosphorus pentoxide (44,122).

Table 8. Boiling Points of Some 1,2,4-Trioxolanes^a

1,2,4-Trioxolanes	CAS Registry Number	Structure	Bp, °C (kPa) ^b
1,2,4-trioxolane	[289-14-5]		18 (2.13)
3,5-dimethyl-1,2,4-trioxolane	[765-57-1]		15 (2.67)
3,3-dimethyl-1,2,4-trioxolane	[22409-33-2]		42–42.5 (18.67)
1,5-dimethyl-6,7,8-trioxabicyclo [3.2.1] octane	[19987-14-5]		58.8 (2.0)

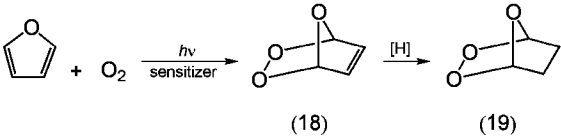
3,5-diphenyl-1,2,4-trioxolane

[23888-15-5]

^a Secondary ozonides (4,44).

^bTo convert kPa to mm Hg, multiply by 7.5.

Cyclic 1,2,4-trioxanes (18 and 19) have been obtained from the photosensitized oxidation of furans (10,44,163). These compounds are 2,3,7-trioxabicyclo [2.2.1] hept-5-ene [6824-18-6] (18) and 2,3,7-trioxabicyclo [2.2.1] heptane [279-56-1] (19).



^b 90 wt % melts at the given temperature.

^c Bp 50 °C at 13.33 kPa^d

^d To convert kPa to mm Hg, multiply by 7.5.

^e Bp 25 °C at 1.6 kPa^d for peroxyacetic acid, and 2.67 kPa^d for peroxypropionic acid.

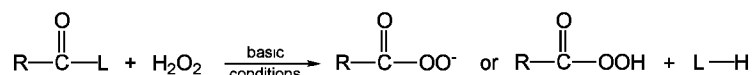
^f Bp 26–29 °C at 1.6 kPa^d.

^g Bp 41–43 °C at 0.067 kPa^d.

Chemical Properties. Organic peroxyacids are not noted for their stability and many lose active oxygen during storage at room temperature. Those that are water soluble hydrolyze slowly to the parent acid and hydrogen peroxide; however, peroxyformic acid hydrolyzes more rapidly. The longer-chain aliphatic members decompose rapidly in methanol. Stabilizers are commonly used for peroxycarboxylic acid solutions, eg, dipicolinic acid, phytic acid, and pyro- and metaphosphates (34,44). Stability of peroxycarboxylic acids increases with increasing molecular weight. The stabilities of peroxybenzoic acids are enhanced when ring substituents are present (34,44).

Hydrolysis is accelerated in the presence of strong acids. However, in the presence of aqueous bases such as sodium hydroxide, the rate of decomposition increases with increasing pH and reaches a maximum at the p*K*_a of the peroxycarboxylic acid (ca 8.25), then decreases at higher pH (169,170). The basic decomposition products include the parent carboxylic acid and singlet oxygen (171,172). Because the maximum rate of decomposition occurs at the p*K*_a, the peroxycarboxylic acid and its anion are involved in the transition state (169).

Peroxycarboxylic acids and precursors to peroxycarboxylic acids are used as bleaches for removal of stains and soils from textiles (173). Precursors to peroxycarboxylic acids are nonperoxidic compounds possessing a reactive acyl group and a good leaving group, L. These precursors react under basic conditions with hydrogen peroxide, inorganic perborates, and inorganic percarbonates to generate peroxycarboxylic acids or salts:



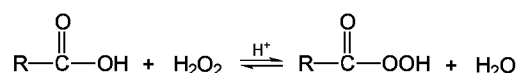
The bleaching activity of peroxycarboxylic acids is less damaging to colors and fibers than hypochlorite bleaches and bleaching can be carried out at temperatures much lower than those required for peroxygen bleaches, such as perborates. Under the basic laundering conditions, oxygen is generated which causes the bleaching action.

Thermal decompositions of peroxycarboxylic acids and their salts can proceed by free-radical and nonradical paths. Often the decomposition products and the rate are affected by the nature of the solvent (44,165,174,175). Peroxycarboxylic acids undergo photodecomposition and radical-induced decomposition (176). They also are decomposed by a variety of metals, metal ions, and complexes (176–180). One-electron transfers occur with transition metals in the same manner as with hydroperoxides. Peroxycarboxylic acids generally are not shock-sensitive but can explode upon heating; relatively pure peroxyformic and peroxyacetic acids are particularly dangerous in this respect (34,44,165). Peroxycarboxylic acids usually are not used as free-radical initiators, although there are many references regarding this application (34).

Peroxycarboxylic acids are among the most powerful organic peroxide oxidizing agents. Typical reactions include epoxidation and hydroxylation of olefins (165) and oxidation of sulfides to sulfoxides or sulfones; disulfides to thiosulfonates; *S*-alkyl thioacetates to alkanesulfonic acids; nitroso compounds or oximes to nitro compounds; nitrosoamines to nitramines; imines to oxaziranes or nitrosoalkane dimers; azo compounds, hydrazones, or amines to azoxy compounds; amines to amine oxides, nitroso compounds, or nitro compounds; azines to azine monoxides or ketones; unsymmetrical dialkylhydrazines to tetraalkyl tetrazenes; aromatic hydrazones to diazo compounds; diazo compounds to ketones; tertiary amines to amine oxides; phenols to muconic acids or quinones; aromatic hydrocarbons to phenols; aromatic ethers to quinones; aldehydes to carboxylic acids; β-diketones to alcohols and α-ketoacids; α-diketones to carboxylic acids; and ketones to esters or lactones, eg, the Baeyer-Villiger reaction (44). Generally, polar solvents accelerate oxidation rates. Reviews on oxidations with magnesium monoperoxyphthalate hexahydrate have been published (181).

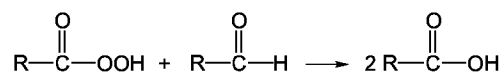
The main industrial uses of peroxycarboxylic acids are in the manufacture of epoxides, synthetic glycerol (qv), and epoxy resins (qv) (165,167,168). They also have been used as disinfectants (177), fungicides, and bleaching agents and for shrink-proofing wool (34).

Synthesis. Many different methods for the preparation of peroxyacids have been described (165). The most widely used method is the direct, acid-catalyzed equilibrium reaction of 30–98 wt % hydrogen peroxide with carboxylic acids (168):



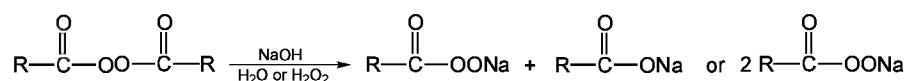
The proportion and concentration of hydrogen peroxide determine the extent of the equilibrium shift. Best results are obtained by using 70–95 wt % hydrogen peroxide. The equilibrium also can be shifted to the right by removing water azeotropically and or under a vacuum. Chelating agents may be added during processing to reduce metal-catalyzed decompositions. When high concentrations of water-soluble peroxycarboxylic acids are not required for an application, the equilibrium mixture can be used. Pure peroxycarboxylic acids can be obtained by careful distillation followed by fractional freezing. Sulfuric acid, methanesulfonic acid, and sulfonic acid ion-exchange resins are the most commonly used acid catalysts. Continuous processes for manufacturing dry solutions of peroxycarboxylic acids free of hydrogen peroxide have been reported (182). Concentrated sulfuric acid has been used with water-insoluble C₂–C₁₁ aliphatic acids. Methanesulfonic acid is used for the higher aliphatic acids and for the aromatic acids.

Another method for producing peroxycarboxylic acids is by autoxidation of aldehydes (168). The reaction is a free-radical chain process, initiated by organic peroxides, uv irradiation, ozone, and various metal salts. It is terminated by free-radical inhibitors (181,183). In certain cases, the peroxycarboxylic acid forms an adduct with the aldehyde from which the peroxycarboxylic acid can be liberated by heating or by acid hydrolysis. If the peroxycarboxylic acid remains in contact with excess aldehyde, a redox disproportionation reaction occurs that forms a carboxylic acid:



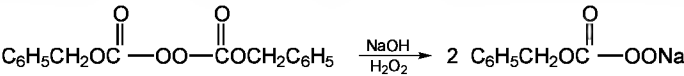
Peroxycarboxylic acids also have been prepared by the reaction of acid chlorides, anhydrides, or boric-carboxylic anhydrides with hydrogen or sodium peroxide. These reactions are carried out at low temperature and with excess peroxide to avoid the formation of diacyl peroxides (44,168,181,184).

Also, basic hydrolysis or perhydrolysis of diacyl peroxides has been used to produce peroxycarboxylic acids (44,181). Perhydrolysis produces two moles of the peroxycarboxylic acid salt:



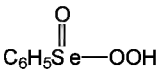
Peroxycarboxylic acids have been obtained from the hydrolysis of stable ozonides with carboxylic acids, perhydrolysis of acylimidazolides, reaction of ketenes with hydrogen peroxide, electrochemical oxidation of alcohols and carboxylic acids, and oxidation of carboxylic acids with oxygen in the presence of ozone (181).

An example of an alkyl monoperoxycarbonic acid, O-benzyl monoperoxycarbonic acid [52123-51-0], was prepared in aqueous methanol solution by basic perhydrolysis of dibenzyl peroxydicarbonate [2144-45-8] and subsequently isolated in 97% purity. It has been used as an epoxidizing agent (185):



Organoperoxysulfonic acids and their salts have been prepared by the reaction of arenesulfonyl chlorides with calcium, silver, or sodium peroxide; treatment of metal salts of organosulfonic acids with hydrogen peroxide; hydrolysis of di(organosulfonyl) peroxides, RS(O)₂—OO—S(O)₂R, with hydrogen peroxide; and sulfoxidation of saturated, nonaromatic hydrocarbons, eg, cyclohexane (44,181).

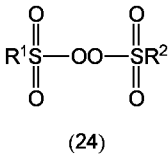
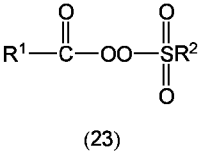
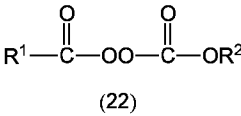
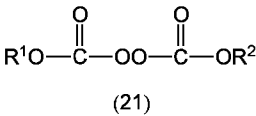
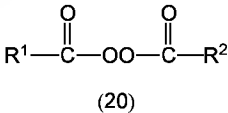
Other Peroxyacids. Benzeneperoxyseleninic acid [62885-97-6], has been



prepared *in situ* from benzeneseleninic acid and hydrogen peroxide and is used to epoxidize terpenic olefins (186) and Baeyer-Villiger oxidation of cyclic ketones.

Acyl Peroxides

The acyl peroxide class is characterized by the following structures:



Acyl peroxides of structure (20) are known as diacyl peroxides. In this structure R¹ and R² are the same or different and can be alkyl, aryl, heterocyclic, imino, amino, or fluoro. Acyl peroxides of structures (21), (22), (23), and (24) are known as dialkyl peroxydicarbonates, OO-acyl O-alkyl monoperoxycarbonates, acyl organosulfonyl peroxides, and di(organosulfonyl) peroxides, respectively. R¹ and R² in these structures are the same or different and generally are alkyl and aryl (4–6,44,166,187,188). Many diacyl peroxides (20) and dialkyl peroxydicarbonates (21) are produced commercially and used in large volumes.

Physical Properties. Almost all liquid diacyl peroxides (20) and concentrated solutions of the solid compounds are unstable to normal ambient temperature storage; many must be stored well below 0°C. Most of the solid compounds are stable at ca 20°C but many are shock-sensitive (187). Other physical constants and properties have been reviewed (187,188). The melting points and refractive indexes of some acyl peroxides are listed in Tables 10–12.

Table 10. Properties of Some Diacyl Peroxides^a

Diacyl peroxide	CAS Registry Number	R group	Mp, °C ^b
diacetyl peroxide	[110-22-5]	<i>Symmetrical, R¹ R²</i> $\text{H}_3\text{C—C}(=\text{O})\text{—OO—C}(=\text{O})\text{—CH}_3$	30
di(chloroacetyl) peroxide	[14295-66-0]	$\text{H}_2\text{CCl—C}(=\text{O})\text{—OO—C}(=\text{O})\text{—ClCH}_2$	85
dipropionyl peroxide	[3248-28-0]	$\text{C}_2\text{H}_5\text{—C}(=\text{O})\text{—OO—C}(=\text{O})\text{—C}_2\text{H}_5$	oil

diisobutyryl peroxide	[3437-84-1]		80–80.5
di(3-carboxypropionyl) peroxide	[123-23-9]		132–133
dipentanoyl peroxide	[925-19-9]		oil
di(4-carboxybutyryl) peroxide	[10195-54-7]		104
di(2-furanylcarbonyl) peroxide	[22023-27-4]		86–87
di(2-thienylcarbonyl) peroxide	[30930-49-5]		92–93
dinicotinoyl peroxide	[13689-05-9]		88–89
diheptanoyl peroxide	[869-90-9]		c
di(cyclohexylcarbonyl) peroxide	[4904-55-6]		oil
dibenzoyl peroxide	[94-36-0]		106–107
di(4-chlorobenzoyl) peroxide	[94-17-7]		137–138
di(4-nitrobenzoyl) peroxide	[1712-84-1]		157–158
di(2-methylbenzoyl) peroxide	[3034-79-5]		54
di(2-carboxybenzoyl) peroxide	[37051-42-6]		156
dioctanoyl peroxide	[762-16-3]		29
di(phenylacetyl) peroxide	[14666-76-3]		41
dinonanoyl peroxide	[762-13-0]		13.0–13.5
di(3,5,5-trimethylhexanoyl) peroxide	[3851-87-4]		d
dicinnamoyl peroxide	[15036-31-4]		133–134
di(benzocyclobutene-4-carbonyl) peroxide	[153213-06-0]		130–132°
didecanoyl peroxide	[762-12-9]		44–45
di(2-naphthaleny carbonyl) peroxide	[38512-20-8]		138–140
didodecanoyl peroxide	[105-74-8]		54.7–55
dihexadecanoyl peroxide	[2697-96-3]		71.4–71.9
dioctadecanoyl peroxide	[3273-75-4]		76.5–76.9
acetyl propionyl peroxide	[13043-82-8]	<i>Unsymmetrical, R¹ = C²H₅; R² as given</i> 	f
acetyl cyclohexyl carbonyl peroxide	[3901-10-8]		42
acetyl benzoyl peroxide	[644-31-5]		37–39
adipoyl bis(acetyl peroxide)	[6039-32-3]		61–62

^a Refs. 44 and 187.

^b Most of these peroxides decompose on melting, some violently.
^c n_D^{20} 1.4340.
^d n_D^{20} 1.4382
^e Ref. 189.
^f n_D^{20} 1.4069

Table 11. Properties of Some Dialkyl Peroxydicarbonates^a

	CAS Registry Number	R ¹	n_D^{20}	Mp, °C ^b
diethyl peroxydicarbonate	[14666-78-5]		1.4065	
di- <i>n</i> -propyl peroxydi-carbonate	[16066-38-9]		1.4091	
diisopropyl peroxydicarbonate	[105-64-6]		1.4034	8–10
dibutyl peroxydicarbonate	[16215-49-9]		1.4129	
di- <i>sec</i> -butyl peroxydicarbonate	[19910-65-7]		1.4112	
dicyclohexyl peroxydicarbonate	[1561-49-5]			46
dibenzyl peroxydicarbonate	[2144-45-8]			101–102
di(2-ethylhexyl) peroxydicarbonate	[16111-62-9]		1.4366	
di(2-phenoxyethyl) peroxydicarbonate	[41935-39-1]			97–100 ^c
di(<i>cis</i> -3,3,5-trimethylcyclohexyl) peroxydicarbonate	[31314-19-9]			78–79 ^d
di(4- <i>tert</i> -butylcyclohexyl) peroxydicarbonate	[15520-11-3]			91–92 ^{e,f}
di(isobornyl) peroxydicarbonate	[40716-59-4]			92–93 ^e
didodecyl peroxydicarbonate	[24356-04-5]			28–30
ditetradecyl peroxydicarbonate	[53220-22-7]			40–42
dihexadecyl peroxydicarbonate	[26322-14-5]			50–53 ^g

^a Refs. 18, 44, and 187.
^b All of the listed peroxides are unstable in the liquid state above ca 20°C; some decompose violently.
^c Ref. 190.
^d Ref. 191.
^e Ref. 192.
^f Ref. 193.
^g Ref. 194.

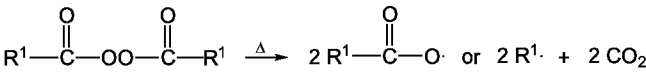
Table 12. Melting Points of Some Organosulfonyl Peroxides^a

Organo-sulfonyl peroxide	CAS Registry Number	Structure	Mp, °C ^b
di(methane-sulfonyl) peroxide	[1001-62-3]	CH ₃ -S(O) ₂ -O- ₂ -	77

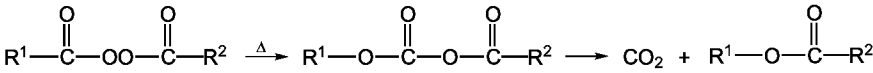
acetyl <i>tert</i> -butane-sulfonyl peroxide	[26305-18-0]		35–37 ^{c, d}
acetyl cyclo-hexane-sulfonyl peroxide	[3179-56-4]		35–36 ^d
acetyl <i>sec</i> -heptane-sulfonyl peroxide	[33970-43-3]		liquid ^e
acetyl (1-methy-cyclo-hexane)-sulfonyl peroxide	[32666-40-3]		liquid ^e
di(benzene-sulfonyl) peroxide	[29342-61-8]	[C ₆ H ₅ -S(O) ₂ -O-] ₂ -	53–54
di(<i>p</i> -toluene-sulfonyl) peroxide	[1886-68-6]	[<i>p</i> -CH ₃ -C ₆ H ₄ -S(O) ₂ -O-] ₂ -	50

^a Refs. 33 and 44.
^b Most of the listed peroxides decompose at mp; some decompose violently.
^c Ref. 195.
^d Ref. 196.
^e Ref. 197.

Chemical Properties. Diacyl peroxides (20) decompose when heated or photolyzed (<300 nm). Although photolytic decompositions generally produce free radicals (198), thermal decompositions can produce nonradical and radical intermediates, depending on diacyl peroxide structure. Symmetrical aliphatic diacyl peroxides of certain structures, ie, diacyl peroxides (20, R¹ = R² = alkyl) without α-branches or with a mono-α-methyl substituent, and diaroyl peroxides (20, R¹ = R² = aryl) thermally decompose almost exclusively by homolysis:



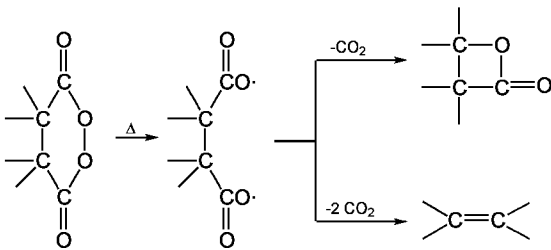
Of these diacyl peroxides the ones that generate the most stable radicals (R¹·) are the most unstable diacyl peroxides. Most other diacyl peroxides decompose by competing free-radical and polar decomposition, ie, carboxy inversion (188). Carboxy inversion occurs to a much greater extent with certain diacyl peroxides having unsymmetrical diacyl peroxide structures (52,187,188,199):



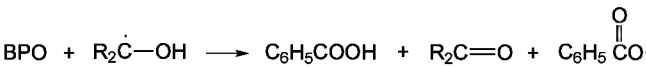
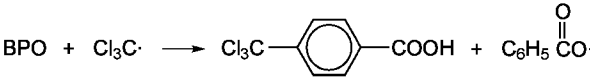
The inversion reaction is more pronounced in polar than in nonpolar solvents and is catalyzed by Lewis acids. For example, under normal conditions, BPO undergoes thermolysis to produce free radicals, whereas in the presence of AlCl₃, BF₃, or SbCl₅, it undergoes carboxy inversion (44). Generally, the diacyl peroxides with electron-donating substituents on one side and electron-withdrawing substituents on the other side of the oxygen–oxygen bond are most susceptible to carboxy inversion.

Diaroyl peroxides and diacyl peroxides without α-branches are significantly more thermally stable than those with mono- or di-α-substituents (188). For example, in 0.2 M benzene, dibenzoyl, diacetyl, dilauroyl, and diisobutyryl peroxides have 10-h HLTs of 73, 69, 62, and 21°C, respectively (72). Based on published rate data in carbon tetrachloride (200), dipivaloyl peroxide has an estimated 10-h HLT of about 5°C. Fluorinated diacyl peroxides are much less stable than nonfluorinated diacyl peroxides, eg, the 10-h HLTs for di(perfluorooctanoyl) peroxide [34434-27-0] and dioctanoyl peroxide are 18 and 51°C, respectively (201). The primary use of most commercial diacyl peroxides (20, R¹ = R² = alkyl or aryl) is initiation of free-radical reactions. In some cases, thermally unstable peroxides, eg, diisobutyryl peroxide, have been prepared *in situ* to initiate vinyl monomer polymerizations (202). Unsymmetrical compositions containing two diacyl peroxide moieties of different thermal stabilities decompose sequentially and can be used as free-radical polymerization initiators for generating peroxypolymers with diacyl peroxide end groups and block copolymers (203,204).

Cyclic diacyl peroxides decompose thermally and photolytically to yield products derived from diradical intermediates (188,198,205) (eq. 31).



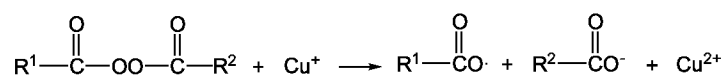
Diacyl peroxides (20, R¹ = R² = alkyl or aryl) also undergo three types of radical induced decomposition (187) all of which produce the radical C₆H₅COO· from BPO. The first type involves direct radical displacement on the oxygen–oxygen bond and is the preferred mode for nucleophilic radicals, eg, ·CH(R)OR'. The second type involves radical addition to, or abstraction from, the hydrocarbyl group adjacent to the peroxide; this is the preferred mode for electrophilic radicals, eg, Cl₃C· (eq. 32). In the last type (eq. 33), there is hydrogen donation from certain hydrogen-donating radicals, eg, ketyls (52,187,188,199).



As a consequence of their susceptibility to radical-induced decomposition, neat and concentrated solutions of diacyl peroxides undergo self-accelerating

decompositions (205). Kinetics studies can be significantly affected if high dilution and free-radical scavengers are not employed. Decomposition rates of diacyl peroxides are faster in more polar solvents.

Diacyl peroxide decompositions also are catalyzed by the metal ions of copper, iron, cobalt, and manganese:



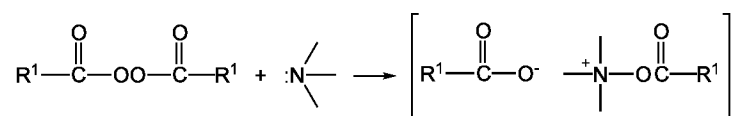
This radical-generating reaction has been used in synthetic applications, eg, aryloxylation of olefins and aromatics, oxidation of alcohols to aldehydes, etc (52,187). Only alkyl radicals, R $\cdot$, are produced from aliphatic diacyl peroxides, since decarboxylation occurs during or very shortly after oxygen–oxygen bond scission in the transition state (187,188,199). For example, diacetyl peroxide is well known as a source of methyl radicals (206).

Hydrolysis and perhydrolysis of diacyl peroxides yields peroxycarboxylic acids. Carbanions react by displacement on oxygen:



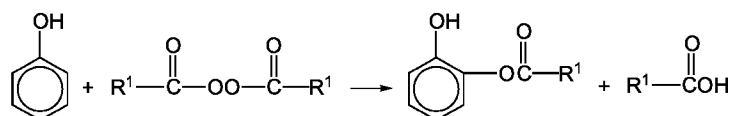
Diacyl peroxides have been reduced with a variety of reducing agents, eg, lithium aluminum hydride, sulfides, phosphites, phosphines, and halide ions (187). Halides yield carboxylic acid salts; (RO) $_3$ P gives acid anhydrides. With iodide ion and certain trivalent phosphorus compounds, the reductions are sufficiently quantitative for analytical purposes.

Amines also react with diacyl peroxides by nucleophilic displacement on the oxygen–oxygen bond forming an ion pair intermediate (187):



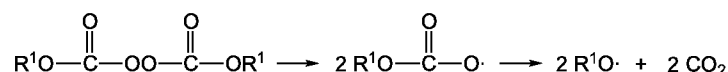
The fate of the ion pair intermediate depends on the structure of the amine and the reaction conditions. Certain tertiary amines, eg, dimethylaniline (DMA), react with specific diacyl peroxides such as dibenzoyl peroxide (BPO) to generate free radicals at ca 20°C. Some reactions, eg, DMA–BPO, are explosive when neat reactants are mixed. Primary and secondary amines do not yield free radicals.

Nonhindered phenols are acyloxyated by diacyl peroxides in nonradical reactions (187):



Phenols with bulky ortho- and para-substituents, eg, phenolic antioxidants, do not undergo this reaction; however, they scavenge radicals generated by thermolysis of diacyl peroxides and other peroxides. Diacyl peroxides react with potassium superoxide, KO $_2$, forming singlet oxygen (207).

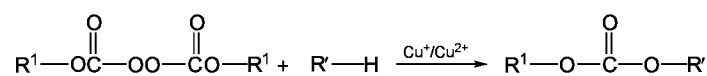
Dialkyl peroxydicarbonates (21) undergo thermolysis to form two alkoxycarbonyloxy radicals that subsequently undergo β -scission to form CO $_2$ and alkoxy radicals:



These low temperature peroxides are susceptible to radical-induced decompositions. This susceptibility largely accounts for the hazards associated with their production and storage. In contrast to diacyl peroxides (20), the true first-order decomposition rates for dialkyl peroxydicarbonates (21) are not affected by the nature of the R group. In free-radical scavenging solvents, eg, trichloroethylene, the decomposition rates of di-*n*-propyl, diisopropyl, di-*sec*-butyl, dicyclohexyl, di-2-ethylhexyl, and dehexadecyl peroxydicarbonates are all essentially the same, ie, 10-h HLT = 49–50°C, whether the R groups are primary, secondary, or cycloalkyl (208). In nonradical-scavenging solvents or environments, decomposition rates are much faster due to competing radical-induced decompositions; such reactions are second-order decompositions (18). Homolysis and induced decomposition reactions of peroxydicarbonates generally proceed at faster rates in solvents of increased polarity.

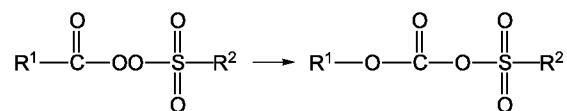
All liquid dialkyl peroxydicarbonates are unstable at ca 20°C and therefore require temperature controlled storage. Many undergo extremely rapid decomposition on warming to ca 20°C, which can be dangerous if the peroxide is confined. Decomposition characteristics, but not necessarily stability, are moderated by dilution. Most solid peroxydicarbonates are stable at 20°C with stability increasing with increasing melting point. Di(2-phenoxyethyl) peroxydicarbonate has a melting point of 97–100°C and is stable in the solid state at 50°C, its 10-h HLT, for at least four weeks (190). However, solid peroxydicarbonates lose their thermal stability when they melt or are dissolved in solvents.

Dialkyl peroxydicarbonates are used primarily as free-radical initiators for vinyl monomer polymerizations (18,208). Dialkyl peroxydicarbonate decompositions are accelerated by certain metals, concentrated sulfuric acid, and amines (44). Violent decompositions can occur with neat or highly concentrated peroxides. As with most peroxides, they liberate iodine from acidified iodides. In the presence of copper ions and suitable substrates, dialkyl peroxydicarbonates have been used to synthesize alkyl carbonates (44):



Thermal decomposition of *OO*-acyl *O*-alkyl monoperoxycarbonates (22, R 1 , R 2 = alkyl or aryl) yield first-order decomposition rates between those of dialkyl peroxydicarbonates (21) and that of the symmetrical diacyl peroxide (20) having the same acyl moiety. These peroxides are also sensitive to radical-induced decompositions (187).

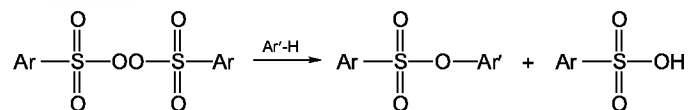
Acyl organosulfonyl peroxides (23) such as acetyl cyclohexanesulfonyl peroxide are efficient radical initiators for vinyl chloride polymerization (24,209). However, in the presence of Lewis acids or strong protic acids most acyl organosulfonyl peroxides decompose in organic solvents by ionic rearrangement with the formation of mixed anhydrides (33):



The rearrangement is self-catalyzed by the organosulfonic acid that is already present or acid from hydrolysis of the mixed anhydride product. If the

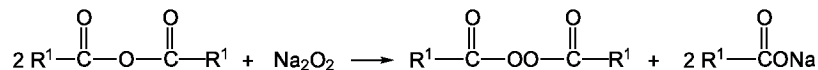
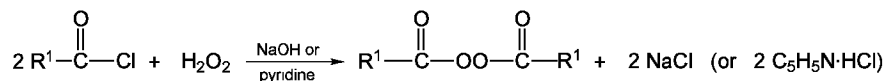
organosulfonic acid is removed, eg, with dry potassium carbonate, the acyl organosulfonyl peroxides decompose predominantly by homolysis to free radicals (33). These peroxides also rapidly decompose in the presence of amines (44).

Di(arenesulfonyl) peroxides (24, $R^1 = R^2 = \text{aryl}$) react with aromatic solvents to form aryl arenesulfonates (33):

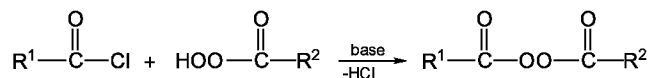


These peroxides also form 1:1 adducts with styrene and form hydrobenzoin diarenesulfonates with stilbenes. Di(benzenesulfonyl) peroxide decomposes in water to phenol and sulfuric acid (33).

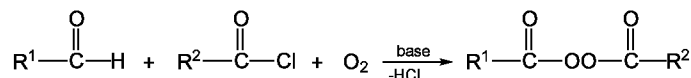
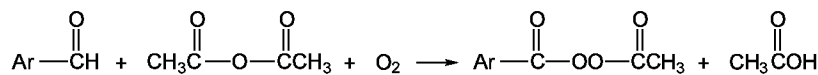
Synthesis. Symmetrical diacyl peroxides (20, $R^1 = R^2 = \text{alkyl or aryl}$) are prepared by the reaction of an acyl chloride or anhydride with sodium peroxide or hydrogen peroxide and a base:



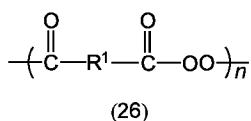
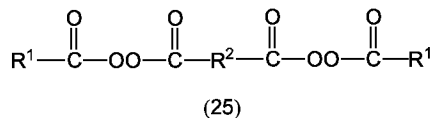
Unsymmetrical diacyl peroxides (20, $R^1 \neq R^2 = \text{alkyl or aryl}$) are prepared by the reaction of acid chlorides or anhydrides with peroxycarboxylic acids in the presence of a base:



The peroxycarboxylic acid can be generated *in situ* by autoxidation of aldehydes, either in the presence of anhydrides or an acyl chloride and a base, eg, sodium carbonate, or basic ion-exchange resins (44,187,188,210):



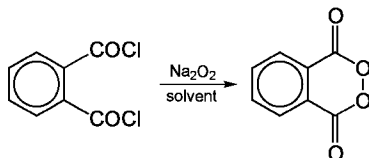
Use of diacid chlorides for acyl chlorides in the latter reaction results in generation of di(diacyl peroxides) (25).



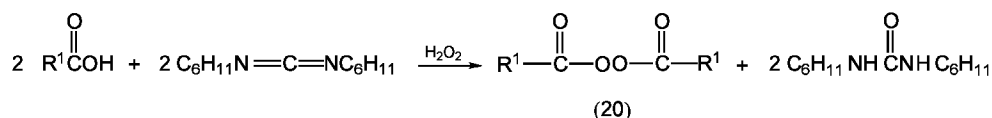
Other unsymmetrical peroxides can be prepared by this reaction by employing other acylating agents, eg, alkyl chloroformates, organosulfonyl chlorides, and carbamoyl chlorides (210). Unsymmetrical and symmetrical di(diacyl peroxides) also are obtained by the reaction of dibasic acid chlorides directly with peroxycarboxylic acids or monoacid chlorides directly with diperoxycarboxylic acids in the presence of a base (44,187,203).

Polymeric diacyl peroxides (26) can be prepared from the reaction of dibasic acid chlorides, eg, succinoyl, fumaryl, sebacoyl, and terephthaloyl chlorides, with sodium or hydrogen peroxide (187).

Cyclic diacyl peroxides can be generated from suitable dibasic acid chlorides and sodium or hydrogen peroxide, especially in dilute solutions (187,205), eg, 2,3-benzodioxin-1,4-dione [4733-52-2] from phthaloyl chloride:



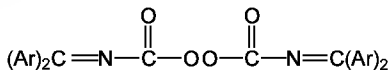
Symmetrical or unsymmetrical diacyl peroxides (20, $R^1, R^2 = \text{alkyl or aryl}$) can be synthesized directly from carboxylic acids and hydrogen peroxide or from peroxycarboxylic acids with dicyclohexylcarbodiimide or *N,N*-dicarbonyldiimidazole as condensing agents (187):



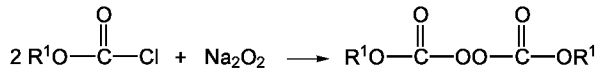
Diacyl peroxides (20, $R^1 = R^2 = \text{alkyl or aryl}$) have been obtained from the oxidation of carboxylic acid potassium salts by Kolbe electrolysis or by elemental fluorine (187).

Fluoroformyl peroxide [692-74-0] (20, $R^1 = R^2 = \text{F}$), has been prepared by the reaction of carbon monoxide, fluorine, and oxygen or by the photolytic reaction of oxalyl fluoride with oxygen (187).

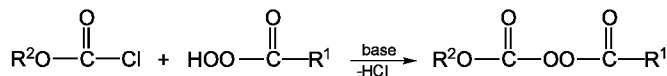
Bis(carbamoyl) peroxides of the following structure were prepared by reaction of the corresponding carbamoyl chlorides and hydrogen peroxide–urea complex in the presence of pyridine (211).



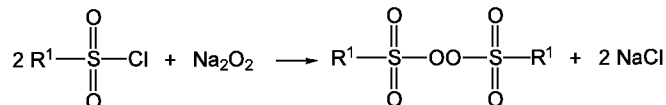
Dialkyl peroxydicarbonates (21) are produced by reaction of alkyl chloroformates with sodium peroxide (44,187):



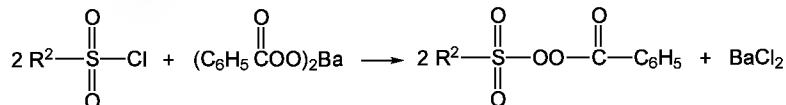
OO-Acyl O-alkyl monoperoxydicarbonates (22) are obtained from the reaction of alkyl chloroformates with peroxycarboxylic acids in the presence of a base (44,212):



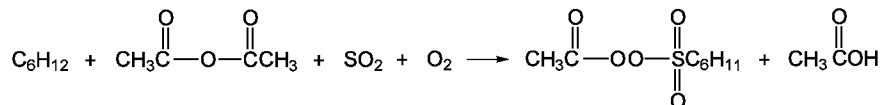
Symmetrical di(organosulfonyl) peroxides (24, $\text{R}^1 = \text{R}^2$) have been prepared by the reaction of organosulfonyl chlorides with sodium peroxide or hydrogen peroxide in the presence of a base (44):



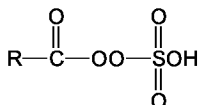
Acyl organosulfonyl peroxides (23) are prepared from the organosulfonyl chlorides and a metal salt of a peroxycarboxylic acid (44):



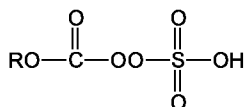
Acetyl cyclohexanesulfonyl peroxide has been produced commercially by the sulfoxidation of cyclohexane, C_6H_{12} , in the presence of acetic anhydride (44):



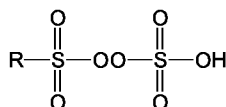
Potassium salts of the peroxides (27–29) are prepared from the reaction of Caro's acid [7722-86-3], H_2SO_5 , with acyl chlorides, chloroformates, or organosulfonyl chlorides in the presence of potassium hydroxide (44).



(27)



(28)



(29)

Alkyl Peroxyesters

Peroxyesters include the alkyl esters of peroxycarboxylic acids; monoperoxydicarboxylic acids; diperoxydicarboxylic acids; monoperoxy- (30) and diperoxydicarboxylic (31) acids; monoperoxy- (32) and diperoxyoxalic (33) acids; peroxycarbamic acids (34); peroxysulfonic acids (35); and peroxyphosphoric acids (36).

Alkyl peroxyesters are commonly named like their nonperoxidic counterparts, except for incorporation of the peroxy- prefix. Trivial names are also commonly used, eg, *tert*-butyl peracetate. Alkyl peroxyesters derived from di- and polybasic peroxyacids use *OO*- or *O*- when required to locate groups, eg, *OO*-*tert*-butyl *O*-isopropyl monoperoxydicarbonate and *OO*-*tert*-butyl *O*-hydrogen monoperoxyoxalate. Descriptions of alkyl peroxyesters have been given in the chemical literature (1,4–6,19,20,44,168,213).

Synthesis. Peroxyesters are prepared by the reaction of alkyl hydroperoxides $\text{R}'\text{OOH}$, with acylating agents, eg, acid chlorides, anhydrides, ketenes, organosulfonyl chlorides, phosgene (qv), alkyl chloroformates, oxalyl chloride, alkyl chlorooxalates, isocyanates, carbamoyl chlorides, carboxylic acids, and esters, under appropriate reaction conditions, according to Figure 2. Reactions with acylating agents that generate hydrogen chloride are carried out in the presence of a base, eg, pyridine or sodium hydroxide, or by using the sodium or potassium salt of the hydroperoxide.

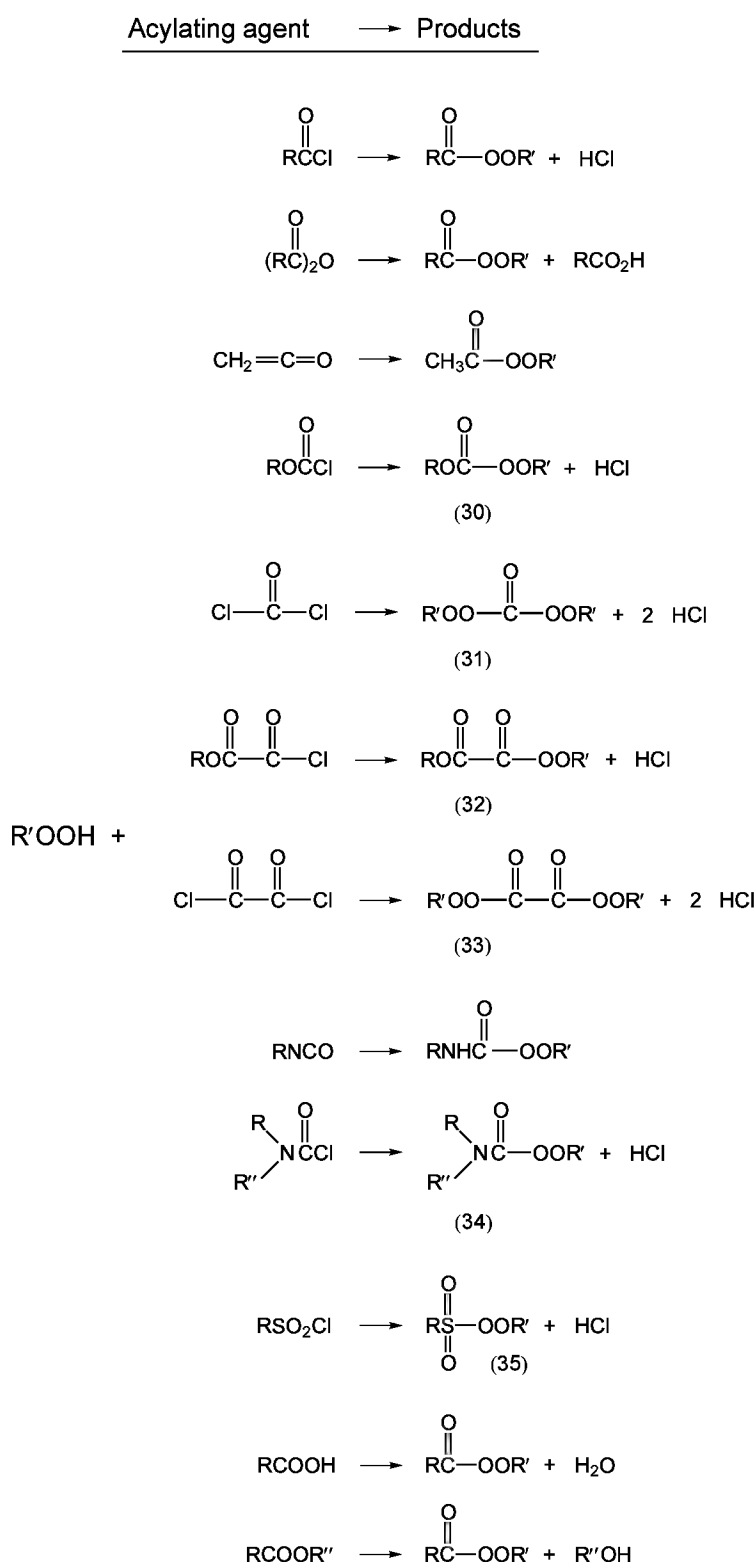


Fig. 2. Synthetic routes to alkyl peroxyesters. The acylating agent reacts with R'OOH in each case.

Because primary and secondary alkyl hydroperoxides are base-sensitive they are converted to peroxyesters by reaction with a ketene or by the reaction of their barium salts with acid chlorides (44).

Peroxyesters may also be prepared by condensation of hydroperoxides with carboxylic acids using condensing agents, eg, dicyclohexylcarbodiimide (214), imidazolides (213), and *p*-toluenesulfonyl chloride with pyridine (213). Suitable esters, eg, monoesters of ethylene glycol, have been used to prepare peroxyesters by ester interchange with alkyl hydroperoxides (215). Generally, reactions of isocyanates to form peroxycarbamates are catalyzed by dibutyltin dilaurate (213,216).

Oligomeric and polymeric peroxides possessing peroxyester moieties have been prepared. Those with peroxyester end groups have been prepared by reaction of the bis(chloroformates) of hydroxy-terminated polymers and oligomers, eg, polycaprolactones, with *tert*-alkyl hydroperoxides in the presence of base (217). Those with pendent peroxyester functions have been synthesized by free-radical copolymerization of *OO-tert*-alkyl *O*-allyl monoperoxycarbonates with vinyl monomers (218). Those possessing peroxyester functions in the polymer backbone have been prepared by base-catalyzed condensation of dihydroperoxides, eg, 2,5-dimethyl-2,5-dihydroperoxyhexane, with dibasic acid chlorides (219) or bis(chloroformates) (220) and by base-catalyzed condensations of hydroxy-hydroperoxides, eg, 3-hydroxy-1,1-dimethylbutyl hydroperoxide, with dibasic acid chlorides or bis(chloroformates) (110). Polymers having peroxycarbamate end groups have been prepared by reaction of isophorone diisocyanate [4098-71-9] with dihydroxy-terminated polyethers and *tert*-butyl hydroperoxide (221).

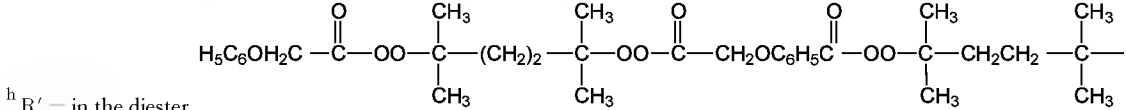
Physical Properties. Properties of some *tert*-alkyl peroxyesters are listed in Table 13 and the properties of some *tert*-alkyl areneperoxysulfonates are given in Table 14. Mass spectra (226), total energies, and dipole moments (227); oxygen–oxygen bond-dissociation energies (44,228); and boiling points, melting points, densities, and refractive indexes (44,168,213) have been reported for a variety of *tert*-butyl peroxycarboxylates.

Table 13. Properties of Some *tert*-Alkyl Peroxyesters^a

Name	CAS Registry Number	R	Mp, °C	Bp, °C (kPa) ^b
<i>R'</i> = <i>tert</i> -butyl(<i>C'</i> H ₃) ₃ <i>C</i> -				
<i>tert</i> -butyl peroxy-formate	[819-50-1]			28.3 (1.33)
<i>tert</i> -butyl peroxy-carbamate	[18389-96-3]		51	
<i>tert</i> -butyl peroxy-acetate	[107-71-1]			22 (0.133)
<i>tert</i> -butyl N,N-di-methyl peroxy-carbamate	[42930-04-1]			43–45 (0.013–0.027)
<i>tert</i> -butyl peroxyiso-butylate	[109-13-7]		45.7	
OO- <i>tert</i> -butyl O-isopropyl mono-peroxy-carbonate	[2372-21-6]			52–55 (0.133)
OO- <i>tert</i> -butyl O-hydrogen mono-peroxy-maleate	[1931-62-0]		114–116	
<i>tert</i> -butyl peroxy-pivalate	[927-07-1]		oil	
di- <i>tert</i> -butyl diperoxy-carbonate	[3236-56-4]			54–55 (0.067)
di- <i>tert</i> -butyl diperoxy-oxalate	[1876-22-8]		50.5–51.5 ^c	
<i>tert</i> -butyl cyclo-hexane peroxy-carboxylate	[20396-49-0]			83–89 (4.0–4.67)
<i>tert</i> -butyl peroxybenzoate	[614-45-9]		8	75–77 (0.267)
<i>tert</i> -butyl 2-ethyl-peroxy-hexanoate	[3006-82-4]		< 30	
<i>tert</i> -butyl 2-phenyl-peroxy-acetate	[3377-89-7]		oil	
<i>tert</i> -butyl 2-carboxy-peroxy-benzoate	[15042-77-0]		104–104.5	
<i>tert</i> -butyl peroxy-nan-oate	[22913-02-6]			52–55 (0.0013)
<i>tert</i> -butyl peroxydeca-n-oate	[16474-36-5]		6.5	
di- <i>tert</i> -butyl diperoxy-adipate	[22158-52-7]		42–43	
<i>tert</i> -butyl 1-naphthalene-peroxy-carboxylate	[13061-74-0]		53–55	
di- <i>tert</i> -butyl diperoxy-phthalate	[2155-71-7]		57.0–57.5	
<i>tert</i> -butyl 2,2-diphenyl-peroxy-acetate	[13144-32-6]		58.2–60	
<i>tert</i> -butyl peroxy-stearate	[2123-89-9]		38.9–39.3	

<i>OO-tert</i> -butyl <i>O-(n</i> -docosyl) mono-peroxy -oxalate	[116753-76-5]	$n\text{-C}_{22}\text{H}_{45}\text{OC}(\text{O})\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\text{C}(\text{CH}_3)_3$	42 ^d
<i>tert</i> -cumyl peroxy-acetate	[34236-39-0]	$\text{CH}_3\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\text{C}(\text{CH}_3)_2\text{C}_6\text{H}_5$	67–68 (0.0067)
<i>tert</i> -cumyl peroxy-pivalate	[23383-59-7]	$t\text{-C}_4\text{H}_9\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\text{C}(\text{CH}_3)_2\text{C}_6\text{H}_5$	18
<i>tert</i> -cumyl peroxy-benzoate	[7074-00-2]	$\text{C}_6\text{H}_5\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\text{C}(\text{CH}_3)_2\text{C}_6\text{H}_5$	45
<i>tert</i> -amyl ^e per-oxyacetate	[690-83-5]	$\text{CH}_3\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-t\text{-C}_5\text{H}_{11}$	65–66 (2.0)
3-hydroxy-1,1 -di-methyl-butyl peroxy-neodecanoate ^f	[95718-78-8]	$\text{H}_{19}\text{C}_9\text{-}t\text{-C}\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\text{C}(\text{CH}_3)_2\text{CH}_2\text{CHOHCH}_3$	oil ^g
2,5-dimethyl- 2,5-di(ben-zo ylperoxy)-hexane ^h	[2618-77-1]	$\text{H}_5\text{C}_6-\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\overset{\text{CH}_3}{\underset{\text{CH}_3}{\text{C}}}-(\text{CH}_2)_2-\overset{\text{CH}_3}{\underset{\text{CH}_3}{\text{C}}}-\text{OO}-\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{C}_6\text{H}_5-\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\overset{\text{CH}_3}{\underset{\text{CH}_3}{\text{C}}}-\text{CH}_2\text{CH}_2-\overset{\text{CH}_3}{\underset{\text{CH}_3}{\text{C}}}-$	118
2,5-dimethyl- 2,5-di(2-phe n-oxyacetyl-p eroxy)-hexane ^h	[6104-83-2]	$\text{H}_5\text{C}_6\text{OH}_2\text{C}-\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\overset{\text{CH}_3}{\underset{\text{CH}_3}{\text{C}}}-(\text{CH}_2)_2-\overset{\text{CH}_3}{\underset{\text{CH}_3}{\text{C}}}-\text{OO}-\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{CH}_2\text{OC}_6\text{H}_5-\overset{\text{O}}{\overset{\parallel}{\text{C}}}-\text{OO}-\overset{\text{CH}_3}{\underset{\text{CH}_3}{\text{C}}}-\text{CH}_2\text{CH}_2-\overset{\text{CH}_3}{\underset{\text{CH}_3}{\text{C}}}-$	71.5–72.5 ¹

^a Refs. 20, 44, and 213.
^b To convert kPa to mm Hg, multiply by 7.5.
^c Ref. 222.
^d Ref. 223.
^e R' = *tert* – amyl (*t*-C₅H₁₁).
^f R' = 3 –hydroxy – 1,1 – dimethylbutyl(CH₃CHOHCH₂C(CH₃)₂–).
^g Ref. 224.



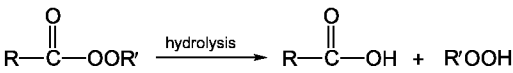
¹ Ref. 225.

Table 14. *tert*-Butyl Areneperoxysulfonates^a

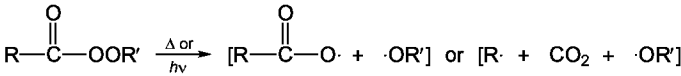
Benzeneperoxy sulfonate	CAS Registry Number	Structure	Mp, °C
<i>tert</i> -butyl benzeneperoxy sulfonate	[18637-19-9]	$\text{C}_6\text{H}_5\overset{\text{O}}{\overset{\parallel}{\text{S}}}-\text{OO}-t\text{-C}_4\text{H}_9$	^b
<i>tert</i> -butyl <i>p</i> -chloro-benzeneperoxy sulfonate	[77482-48-5]	$p\text{-Cl-C}_6\text{H}_4\overset{\text{O}}{\overset{\parallel}{\text{S}}}-\text{OO}-t\text{-C}_4\text{H}_9$	30–35 ^c
<i>tert</i> -butyl <i>p</i> -methyl-benzeneperoxy sulfonate	[77482-49-6]	$p\text{-CH}_3\text{-C}_6\text{H}_4\overset{\text{O}}{\overset{\parallel}{\text{S}}}-\text{OO}-t\text{-C}_4\text{H}_9$	36.5–37
<i>tert</i> -butyl <i>p</i> -methoxy-benzeneperoxy sulfonate	[77482-50-9]	$p\text{-CH}_3\text{O-C}_6\text{H}_4\overset{\text{O}}{\overset{\parallel}{\text{S}}}-\text{OO}-t\text{-C}_4\text{H}_9$	47 ^c

^a Refs. 33 and 44.
^b *n*_D²⁵ 1.4629.
^c Explodes at mp.

Chemical Properties. Alkyl peroxyesters are hydrolyzed more readily than the analogous nonperoxidic esters and yield the original acids and hydroperoxides from which they were prepared rather than alcohols and peroxyacids:



The *tert*-alkyl peroxyesters undergo homolysis, thermally and photochemically, to generate free radicals (168,213,229–232):



Simultaneous cleavage of two bonds occurs when R· is relatively stable (168) or when R is bulky. Thermal first-order decomposition rates of *tert*-alkyl peroxyesters are influenced strongly by structure. Variation of both R and R' groups of peroxyesters provides a convenient means of altering the relative thermal activity of peroxyesters. For example, increasing the steric bulk of either or both of R and R' generally lowers the thermal stability of a peroxyester. These trends are illustrated in Table 15. *tert*-Butyl peroxyacetate is significantly more thermally stable and less active than 3-hydroxy-1,1-dimethylbutyl peroxyneodecanoate. Although other factors affect thermal stability, the concept of steric bulk of R and R' can be used to qualitatively predict peroxyester reactivity trends.

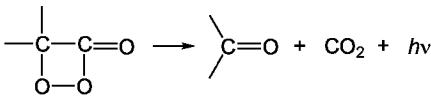
Table 15. Effect of *tert*-Alkyl Peroxyester Structure on 10-h HLT, R-C(O)-OO-R'^a

R	CAS Registry Number	10-h HLT, °C (solvent)
C H ₅ —C(O)—OO— <i>t</i> -C ₄ H ₉	[614-45-9]	104 (dodecane)
CH ₃ —C(O)—OO— <i>t</i> -C ₄ H ₉	[107-71-1]	102 (decane)
(CH ₃) ₂ CHO—C(O)—OO— <i>t</i> -C ₄ H ₉	[2372-21-6]	99 (benzene)
(CH ₃) ₂ CH—C(O)—OO— <i>t</i> -C ₄ H ₉	[109-13-7]	82 (dodecane)
(CH ₃) ₃ C—C(O)—OO— <i>t</i> -C ₄ H ₉	[927-07-1]	58 (TCE) ^b
C H ₅ CH ₂ —C(O)—OO— <i>t</i> -C ₄ H ₉	[3377-89-7]	66 (chlorobenzene) ^c
(C H ₅) ₂ CH—C(O)—OO— <i>t</i> -C ₄ H ₉	[13144-32-6]	37 (cumene) ^d
(C H ₅) ₃ C—C(O)—OO— <i>t</i> -C ₄ H ₉	[10357-71-8]	11 (cumene) ^e
<i>t</i> -C ₉ H ₁₉ —C(O)—OO— <i>t</i> -C ₄ H ₉	[26748-41-4]	48 (TCE)
<i>t</i> -C ₉ H ₁₉ —C(O)—OO— <i>t</i> -C ₄ H ₉	[68299-16-1]	46 (TCE)
<i>t</i> -C ₉ H ₁₉ —C(O)—OO— <i>t</i> -C ₄ H ₉ ^f	[51240-95-0]	44 (TCE)
<i>t</i> -C ₉ H ₁₉ —C(O)—OO— <i>t</i> -C ₄ H ₉	[26748-47-0]	38 (TCE)
<i>t</i> -C ₉ H ₁₉ —C(O)—OO— <i>t</i> -C ₄ H ₉	[95718-78-8]	37 (TCE)
<i>t</i> -C ₄ H ₉ —OO—C(O)—C(O)—OO— <i>t</i> -C ₄ H ₉	[1876-22-8]	26 (benzene) ^g

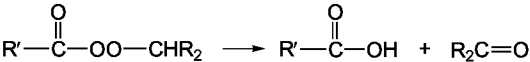
^a From or calculated from Ref. 22 unless otherwise noted.
^b Trichloroethylene.
^c Ref. 233.
^d Ref. 234.
^e Ref. 235.
^f 1,1,3,3-Tetramethylbutyl.
^g Ref. 222.

Table 15 shows that peroxyester stability decreases for the *tert*-alkyl groups in the following order: *tert* - butyl > *tert* - amyl > *tert* - octyl > *tert* - cumyl > 3 - hydroxy - 1, 1 - dimethylbutyl. The order of activity of the R' group in peroxyesters is also observed in other *tert*-alkyl peroxides. Peroxyesters derived from benzoic acids and non-αbranched carboxylic acids are more stable than those derived from mono-α-branched acids which are more stable than those derived from di-α-branched acids (19,21,168). The size of the α-branch also is important, since steric acceleration of homolysis occurs with increasing branch size (236). Suitably substituted peroxyesters show rate enhancements because of anchimeric assistance (168,213,237).

β-Peroxlactones undergo thermal decarboxylation to carbonyl compounds by the initial formation of a 1,5-diradical (238). α-Peroxlactones undergo similar decarboxylation, emitting light since the ketone is generated in the triplet excited state (85,239,240):

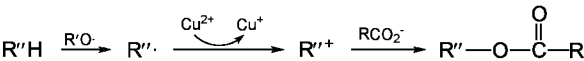


Primary and secondary alkyl peroxyesters thermally decompose by a nonradical process, giving almost quantitative yields of carboxylic acids and carbonyl compounds (213,241):



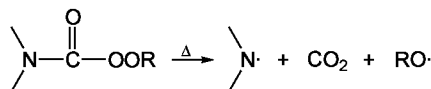
tert-Alkyl peroxyesters are much less sensitive to radical-induced decompositions than diacyl peroxides. Induced decomposition is only significant in peroxyesters containing nonhindered α-hydrogens or α, β-unsaturation (213,242).

Peroxyesters decompose by an electron-transfer process catalyzed by transition metals (44,168,213) (eq. 34). This reaction has been used synthetically to bond an acyloxy group to appropriate coreactive substrates (eq. 35).



Apparently the alkoxy radical, $R'O\cdot$, abstracts a hydrogen from the substrate, $R''-H$, and the resulting radical, $R''\cdot$, is oxidized by Cu^{2+} (one-electron transfer) to form a carbonium ion that reacts with the carboxylate ion, RCO_2^- . The overall process is a chain reaction in which copper ion cycles between $+1$ and $+2$ oxidation states. Suitable substrates include olefins, alcohols, mercaptans, ethers, dienes, sulfides, amines, amides, and various active methylene compounds (44). This reaction can also be used with *tert*-butyl peroxy-carbamates to introduce carbamoyloxy groups to these substrates (243).

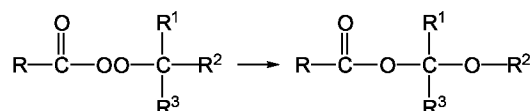
Normally, *tert*-alkyl peroxy-carbamates undergo thermolysis to yield free radicals and carbon dioxide (20).



The first-order decomposition rates of *tert*-alkyl peroxy-carbamates are strongly influenced by structure, eg, electron-donating substituents on nitrogen increase the rate of decomposition, and some substituents increase sensitivity to induced decomposition (20). *tert*-Alkyl peroxy-carbamates have been used to initiate vinyl monomer polymerizations and to cure rubbers (244). They liberate iodine quantitatively from hydriodic acid solutions. Decomposition products include carbon dioxide, hydrazo and azo compounds, amines, imines, and *O*-alkylhydroxylamines. Many peroxy-carbamates are stable at ca 20°C but decompose rapidly and sometimes violently above 80°C (20,44).

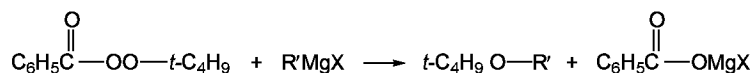
Acid-catalyzed, ionic decompositions have been reported for peroxyesters, $RC(O)-OOR'$, in which the R group can form a particularly stable carbonium ion, eg, tropylium ion (213).

Criegee rearrangement competes with homolysis in *tert*-alkyl peroxyesters, $RC(O)-OOCR^1R^2R^3$, in which R is strongly electron-withdrawing and the *tert*-alkyl group, ie, $CR^1R^2R^3$, contains a group with high migratory aptitude and ability to stabilize adjacent carbonium ions (213). The rearrangement converts the peroxyester to a nonperoxidic ester:



Some peroxyesters are difficult to prepare because of this facile rearrangement, eg, attempts to prepare triphenylmethyl peroxyesters yield only rearrangement products (44).

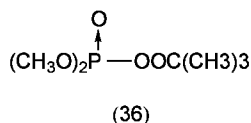
tert-Butyl peroxybenzoate and probably other *tert*-alkyl peroxyesters react with carbanions, eg, Grignard reagents or phenyllithium, at low temperatures to form ethers, via nucleophilic displacement at the oxygen adjacent to the *tert*-alkyl group (44):



Lithium amides of primary *tert*-alkylamines yield *N*-(*tert*-alkyl)-*O*-(*tert*-butyl)hydroxylamines, whereas lithium amides of primary alkylamines yield *N*-alkylbenzamides and $LiOO-t-C_4H_9$ due to nucleophilic attack on the carbonyl group (245).

The instability of *tert*-butyl areneperoxysulfonates is increased by the presence of electron-withdrawing substituents on the aromatic ring and decreased by electron-donating substituents. However, even the most stable members decompose violently on warming, as indicated in Table 14. These peroxyesters appear to decompose heterolytically without the formation of free radicals (44).

OO-tert-Alkyl dialkyl monoperoxyphosphates such as *OO-tert*-butyl dimethyl monoperoxyphosphate [*18963-64-9*] (mp, 23–25°C) (36) have been prepared and appear to decompose heterolytically at the oxygen–oxygen bond (33).



The use of monomers that do not homopolymerize, eg, maleic anhydride and dialkyl maleates, reduces the shock sensitivity of *tert*-butyl peroxyesters and other organic peroxides, presumably by acting as radical scavengers, that prevent self-accelerating, induced decomposition (246).

The main industrial use of *tert*-alkyl peroxyesters is in the initiation of free-radical chain reactions, primarily for vinyl monomer polymerizations. Decomposition of unsymmetrical diperoxyesters, in which the two peroxyester functions decompose at different rates, results in the formation of polymers of enhanced molecular weights, presumably due to chain extension by sequential initiation (204).

Manufacture and Processing

The general chemical approaches for syntheses of commercial organic peroxides have been described in earlier sections of this article. Owing to the inherent hazards of organic peroxides, they are almost never distilled or confined during manufacture. Generally, open reactors are employed that can be easily vented and deluged with water if an unanticipated exotherm occurs. The preferred materials of reactor construction are 316 stainless steel, plastic, and glass. Brass and copper should be avoided owing to the sensitivity of many organic peroxides to copper salts. Significant cooling capacity is required to handle reaction exotherms and to maintain temperature. Because over 100 different organic peroxides are produced commercially, organic peroxide producers manufacture many organic peroxides in the same equipment. Batch processing is generally employed when relatively small production volumes are required, whereas semicontinuous and continuous processing are employed when larger production volumes are required and when safety is a primary issue (247). In batch processing, a reactor is charged with all of the reactants, optional solvents, and catalysts early in the process and the reaction is allowed to proceed. The reaction is completed under controlled conditions and the entire batch of product is purified with various aqueous washes and extractions and dried using a desiccant, dry gas blowing, or stripping under a vacuum. In continuous processing, the raw materials in relatively small amounts are continuously added to a mixing zone and held in the mixing zone for the required reaction time. Centrifuges are often used for purification. Drying is accomplished as previously described. Semicontinuous processing employs some continuous processing steps, usually the reaction step, and some batch processing steps. Continuous processes are significantly safer to operate than batch processes as smaller amounts of organic peroxides are continuously in process. Besides safer continuous processing, another trend has been the use of reactants of higher purity. These process improvements have resulted in reduced environmental impact as unplanned process decompositions have been decreased and waste streams have been reduced.

Economic Aspects

Prices of commercial organic peroxides range from ca \$2.50 to >\$35/kg, depending on peroxide type, production volume, assay, nature of formulation, cost of raw materials, and degree of special processing and handling requirements. For example, >98 wt% BPO is sold only in 0.45-kg paper-lined bags for >\$13/kg, whereas the more stable 78 wt% assay, wetted product is sold in 11.3-kg quantities for ca \$8/kg, 100 wt% basis. Prices of low temperature peroxides, eg, the liquid peroxydicarbonates and low temperature peroxyesters, are high, ie, >11/kg, 100 wt% basis, because they must be refrigerated during storage and shipment.

Analytical and Test Methods

Analytical methods for organic peroxides have been reviewed (248). The most general chemical analytical methods for organic peroxides involve the reduction of the peroxide group followed by the determination of the excess reducing agent or of the oxidized form of the reducing agent (249–252). The approximate order of organic peroxides in terms of decreasing ease of reduction, based on polarographic studies and reduction with iodide ion, is peroxyacids > diacyl peroxides > hydroperoxides > peroxydicarbonates > diperoxyketals > ozonides > peroxyesters > dialkyl peroxides. There is no one general method of analysis for all peroxides because of the various reactivity differences between classes and within each class.

The most commonly used reducing agent is iodide ion:



The liberated iodine, as the complex triiodide ion, may be titrated with standard thiosulfate solution. A general iodometric assay method for organic peroxides has been published (253). Some peroxyesters may be determined by ferric ion-catalyzed iodometric analysis or by cupric ion catalysis. The latter has become an ASTM Standard procedure (254). Other reducing agents are ferrous, titanous, chromous, stannous, and arsenite ions; triphenylphosphine; diphenyl sulfide; and triphenylarsine (255,256).

Polarography is a useful instrumental technique for analysis of peroxides that are irreversibly reduced at the mercury electrode, eg, hydroperoxides, peroxyesters, diacyl peroxides, and peroxydicarbonates (257–259). Examination of ir absorption bands in the carbonyl region and at 800–900 cm⁻¹ is useful for analysis of diacyl peroxides, peroxyesters, and peroxydicarbonates (260). Gas chromatography (gc) may be used to analyze thermally stable hydroperoxides, peroxyesters, diacyl peroxides, dialkyl peroxides, and their mixtures (261–263). Mass spectrometry, gc-mass spectrometry, nmr, and infrared and uv-absorption techniques are useful for structural characterization and analysis. Thin-layer, paper, and column chromatography are useful for identification and separation of peroxides in complex mixtures (264–267). High pressure liquid chromatography is excellent for the separation and analysis of peroxide mixtures, especially those containing components with low volatility or temperature sensitivity. A liquid chromatographic method for assaying dicumyl peroxide has been published (268).

There are many colorimetric methods used for trace analysis of peroxides using reagents such as ferrous ion, leuco base of methylene blue, *sym*-diphenylcarbohydrazide, titanium(IV), iodide ion, and *N,N'*-dimethyl-*p*-phenylenediamine. The latter two are the most commonly used reagents (269,270).

Health and Safety Factors

Toxicology. In general, organic peroxides are characterized by a low order of acute toxicity. Most organic peroxides have some oxidizing properties and are irritants. Hydroperoxides, peroxyacids, ketone peroxides, and other peroxides with hydroperoxy groups generally are more irritating to the skin and eyes than peroxyesters, peroxydicarbonates, diacyl peroxides, or dialkyl peroxides. Hydroperoxides, MEKPs, and some other ketone peroxides are particularly injurious to eyes and can cause blindness (271,272). In contrast, most peroxyesters seem to be more irritating to the skin than the eye. Of the organic peroxides that have been evaluated for skin sensitization, only BPO has produced positive effects under specific conditions. Most of the available toxicity data on commercial organic peroxides are summarized in the literature (273,274).

Two studies conducted by the National Toxicology Program have suggested that effects after repeated administration are primarily due to the irritant nature of these materials. *tert*-Butyl peroxybenzoate, administered by gavage to rats and mice for up to 90 days, produced damage in the stomach with no systemic toxicity (275). A 45% formulation of MEKP in dimethyl phthalate applied to the skin of rats and mice for 13 weeks produced skin corrosion which was limited to the application site. Changes in the spleen and bone marrow were considered secondary responses to the ulcerative skin lesions (276). Occupational exposure to dicumyl peroxide in the atmosphere has been reported to produce nasal irritation and crusting with the appearance of visible blood vessels in the nose. These effects have also been reproduced in rabbits exposed by inhalation or by direct installation of this peroxide in the nose. Rats exposed to α -cumyl hydroperoxide by inhalation for 90 days had signs of irritation of the respiratory tract and mucous membranes (including eyes and noses) at 124 mg m⁻³. Exposures were terminated after five days due to excessive toxicity. No adverse effects were reported at 31 mg m⁻³ (277).

There is limited evidence to suggest that organic peroxides are carcinogenic. Generally, most studies have focused on the skin tumor promoting potential of these materials. For example, BPO does not seem to be a complete carcinogen, but can act as a promoter of skin tumors in mice treated with known initiating carcinogens. Similar findings have been reported for lauroyl peroxide and peroxybenzoic acid (278).

Several organic peroxides have been tested for mutagenicity in bacterial and animal cells, and in animals. In general, a variety of short-term screening studies have indicated that many organic peroxides are mutagenic when tested in bacterial and animal cells, but are not mutagenic when tested in whole animal studies. For example, *tert*-butyl hydroperoxide was mutagenic in the Ames test in mouse lymphoma assay, and produced chromosomal damage in Chinese hamster ovary cells, but did not demonstrate mutagenic activity in rat bone marrow after inhalation exposure for five days, and in a mouse dominant lethal study (274). Similar findings have been reported with MEKPs and *tert*-butyl peroxybenzoate (277,278).

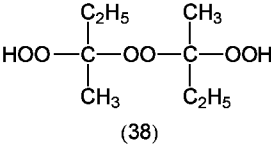
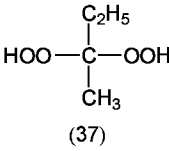
Decomposition Hazards. The main causes of unintended decompositions of organic peroxides are heat energy from heating sources and mechanical shock, eg, impact or friction. In addition, certain contaminants, ie, metal salts, amines, acids, and bases, initiate or accelerate organic peroxide decompositions at temperatures at which the peroxide is normally stable. These reactions also liberate heat, thus further accelerating the decomposition. Commercial products often contain diluents that desensitize neat peroxides to these hazards. Commercial organic peroxide decompositions are low order deflagrations rather than detonations (279).

Methyl and ethyl hydroperoxides explode violently with heating or jarring, and their barium salts are extremely explosive in the dry state. The alkylidene peroxides that are derived from hydrogen peroxide and low molecular weight aldehydes or ketones are sensitive and explode violently. Polymeric peroxides are often sensitive, eg, the polymeric peroxide of dimethylketene [67772-28-5] explodes in the dry state when rubbed at -80°C. Owing to sensitivity and high energy potential, pure low molecular weight peroxyacids, ether peroxides, and diacetyl, dimethyl, and dipropionyl peroxides should be handled only in small quantities and with extreme care.

Within each organic peroxide class, sensitivity to shock increases with increasing active oxygen content. Besides active oxygen content, sensitivity is also dependent on the heat of decomposition, the activation energy, and the decomposition kinetics. Some peroxides that usually are regarded as relatively nonhazardous under certain conditions may be highly hazardous under other conditions. For instance, α -cumyl hydroperoxide is quite thermally stable at room temperature. However, α -cumyl hydroperoxide decomposes violently when inadvertently contaminated with strong acids. In addition, many mixtures of organic compounds and aqueous hydrogen peroxide detonate upon initiation (280).

The organic peroxides and peroxide compositions produced commercially are those that can be manufactured, shipped, stored, and used safely. Organic peroxides can be thermally and mechanically desensitized by wetting or by dilution with suitable solvents, inert solid fillers, or insoluble liquids (suspension of solid peroxides in liquid plasticizers or water, and emulsions of liquid peroxides in water).

Dry BPO is hazardous. Desensitized BPO is commercially available in several solid forms, eg, wetted solid BPO compositions, wetted BPO pastes, and aqueous BPO dispersions. Because of the heat-absorbing capacity of interstitial water, wetted formulations are much more difficult to decompose or burn than comparable dry compositions having the same BPO concentration. The presence of small amounts of water in BPO significantly reduces its burning hazards. Commercial MEKP compositions contain two principal MEKP structures, (37) and (38), small amounts of free hydrogen peroxide, plasticizer solvents (commonly phthalates), and other solvents. Shipping of ketone peroxide solutions having more than 9 wt % active oxygen is unlawful.



Recommendations for safe handling and storage of commercial organic peroxides are available from organic peroxide manufacturers (25,26,281). Test methods have been developed for evaluating the hazards associated with the handling, storage, and transportation of organic peroxides, particularly commercial organic peroxide compositions. Many test methods are used to determine the labeling requirements for commercial organic peroxide compositions. Table 16 lists the important hazard tests used with commercial organic peroxides, references that describe the test methodologies and interpretation of results, and the relevance of the hazard testing with respect to requirements for handling, storage, shipping, and labeling.

Table 16. Hazard Tests for Organic Peroxides^a

Hazard test	Test relevant to: ^b
impact sensitivity	H
burning test	H,S
SETA flash point ^c	H,S,L
pressure vessel test	H,S,T
rapid heat test	H
self-accelerating decomposition temperature (SADT) test	H,S,T,L
modified trauzl block test	H,S,T
thermal stability test	H,S,T
bulk vent tests ^d	H,S,T

^a Ref. 282, unless otherwise noted.
^b H handling, S storage, T transportation, and L labeling.
^c Ref. 283.
^d Refs. 284 and 285.

In thermal stability tests, small samples of the peroxide composition are tested for various periods at several temperatures. The loss of activity is a guide in setting the maximum allowable storage and shipping temperature for the commercial organic peroxide composition. The self-accelerating decomposition temperature (SADT) test is a refinement of the thermal stability test and is used to determine the maximum allowable storage and transportation temperature for an organic peroxide composition in a particular package size. The SADT is the lowest ambient (storage) temperature at which the product undergoes self-accelerating decomposition within seven days. Generally, the SADT decreases as the package size increases. Thus, the SADT is usually determined for a peroxide formulation in its largest commercial package. Because large-scale testing can destroy the test ovens, attempts have been made to develop smaller scale tests that would be less destructive and consume less product. The results of such small-scale tests have been used to reliably determine the SADT of a product in any type and size of package. Examples of small-scale tests include the adiabatic storage test (AST) (286), the isothermal storage test (IST) (287), and the Wärmestau Verfahren test (287,288). An attempt was made to use an accelerating rate calorimeter (ARC) for determining SADTs of peroxides (289).

A hazard classification system has been developed for organic peroxides, particularly commercial organic peroxide compositions. A numeric rating is assigned based on results of several tests including the pressure vessel test, the rapid heat test, the SADT test, the impact sensitivity test, the modified trauzl block test, and the burning test (282).

To meet customer needs, organic peroxide producers have developed highly diluted solutions of organic peroxides that can be shipped in large bulk containers or tank trucks. The need for safe bulk shipping led to development of large-scale vent tests that determine the size and rating of venting devices for bulk containers. Vent testing of organic peroxide compositions was initially done on small-volume vessels (ca 20 mL) (72) and subsequently on larger volume vessels (121 and 10,000 L) (284,285).

In 1984 the United Nations (UN) Committee on the Transportation of Dangerous Goods, made up of experts from Prins Maurits Laboratory (TNO), Bundesanstalt für Materialprüfung (BAM), and the organic peroxide producers, developed a test procedure for the classification of organic peroxide compositions for transport purposes. The test procedure was accepted by most of the industrial countries of the world. The Department of Transportation (DOT) mandated that the United States peroxide industry would comply with the UN classification system by October 1993. The test scheme is used to assign each organic peroxide composition to one of seven hazard categories, A through G, with A being the most hazardous and G being exempt from regulations. The flow chart classification scheme for organic peroxides has been published (290).

Material Safety Data Sheets (MSDS) and the organic peroxides producers' recommendations should be followed carefully for handling and storage of organic peroxide compositions.

Uses

There are more than 100 commercially available organic peroxides in well over 300 formulations, eg, neat liquids and solids, and pastes, powders, solutions, dispersions, and emulsions, that have utility in many commercial applications (13,14,16,21,22,24–26,44,98,99,208,209,291–305). Many of the commercially available peroxides are listed in Table 17 along with 10-h HLTs.

Table 17. 10-h HLTs of Commercial Organic Peroxides^a

Organic peroxide	CAS Registry Number	10-h HLT, ^b °C	Reference
diisobutryl peroxide	[3437-84-1]	21 (B); 25 (TCE)	72
3-hydroxy-1,1-dimethylbutyl peroxyneodecanoate	[95718-78-8]	37 (TCE)	22
tert-cumylperoxyneodecanoate	[26748-47-0]	38 (TCE)	22
OO-tert-butyl O-docosyl monoperoxyoxalate	[116753-76-5]	38	223

acetyl cyclohexanesulfonyl peroxide	[3179-56-4]	31 (B); 40 (TCE)	209
3-hydroxy-1,1-dimethylbutyl peroxyneoheptanoate	[110972-57-1]	41 (AMS)	22
<i>tert</i> -cumyl peroxyneoheptanoate	[104852-44-0]	43 (TCE)	22
<i>tert</i> -amyl peroxyneodecanoate	[68299-16-1]	46 (TCE)	22
<i>tert</i> -butyl peroxyneodecanoate	[26748-41-4]	48 (TCE)	22
di(2-ethylhexyl) peroxydicarbonate	[16111-62-9]	49 (TCE)	22
diisopropyl peroxydicarbonate	[105-64-6]	50 (TCE)	72
di(<i>sec</i> -butyl) peroxydicarbonate	[19910-65-7]	50 (TCE)	22
<i>tert</i> -amyl peroxy-pivalate	[29240-17-3]	55 (TCE)	22
di(2,4-dichlorobenzoyl) peroxide	[133-14-2]	54 (B)	72
<i>tert</i> -butyl peroxy-pivalate	[927-07-1]	58 (TCE)	22
di(3,5,5-trimethylhexanoyl) peroxide	[3851-87-4]	59 (B); 61 (TCE)	72,22
didecanoyl peroxide	[762-12-9]	61 (B); 65 (TCE)	72,22
didodecanoyl peroxide	[105-74-8]	62 (B); 64 (TCE)	72,22
2,5-dimethyl-2,5-di(2-ethylhexanoylperoxy) hexane	[13052-09-0]	73 (DC)	22
diacetyl peroxide	[110-22-5]	69 (B)	72
<i>tert</i> -amyl 2-ethylperoxyhexanoate	[686-31-7]	75 (D)	22
<i>tert</i> -butyl 2-ethylperoxyhexanoate	[3006-82-4]	77 (D)	22
dibenzoyl peroxide (BPO)	[94-36-0]	73 (B)	22
<i>tert</i> -butyl peroxyisobutyrate	[109-13-7]	82 (D)	22
1,1-di(<i>tert</i> -amylperoxy)cyclohexane	[15667-10-4]	93 (D)	22
1,1-di(<i>tert</i> -butylperoxy)-3,3,5-trimethylcyclohexane	[6731-36-8]	96 (D)	22
1,1-di(<i>tert</i> -butylperoxy)cyclohexane	[3006-86-8]	97 (D)	22
<i>OO-tert</i> -butyl <i>O</i> -hydrogen monoperoxymaleate	[1931-62-0]	87 (AC)	22
<i>OO-tert</i> -butyl <i>O</i> -isopropyl monoperoxycarbonate	[2372-21-6]	99 (B)	22
<i>OO-tert</i> -butyl <i>O</i> -(2-ethylhexyl) monoperoxycarbonate	[34443-12-4]	100 (D)	22
2,5-di(benzoylperoxy)-2,5-dimethylhexane	[2618-77-1]	100 (B)	22
<i>tert</i> -butyl peroxy-3,5,5-trimethylhexanoate	[13122-18-4]	102 (B)	72
<i>tert</i> -amyl peroxyacetate	[690-83-5]	100 (D)	22
<i>tert</i> -amyl peroxybenzoate	[4511-39-1]	100 (D)	22
<i>tert</i> -butyl peroxyacetate	[107-71-1]	102 (DC)	22
<i>tert</i> -butyl peroxybenzoate	[614-45-9]	104 (D)	22
2,2-di(<i>tert</i> -butylperoxy)butane	[2167-23-9]	107 (D)	22
2,2-di(<i>tert</i> -amylperoxy)propane	[3052-70-8]	108 (D)	22
ethyl 3,3-di-(<i>tert</i> -amylperoxy)butyrate	[67567-23-1]	112 (D)	22
ethyl 3,3-di-(<i>tert</i> -butylperoxy)butyrate	[55794-20-2]	114 (D)	22
dicumyl peroxide	[80-43-3]	117 (DC)	22
1,3-di[α -(<i>tert</i> -butylperoxy)isopropyl]-benzene	[25155-25-3]	119 (D)	22
2,5-di(<i>tert</i> -butylperoxy)-2,5-dimethylhexane	[78-63-7]	120 (D)	22
di- <i>tert</i> -amyl peroxide	[10508-09-5]	123 (D)	22
di- <i>tert</i> -butylperoxide	[110-05-4]	129 (DC)	22
2,5-di(<i>tert</i> -butylperoxy)-2,5-dimethyl-3-hexyne	[1068-27-5]	131 (D)	22
<i>p</i> -menthane hydroperoxide	[26762-92-5]	133 (B) ^c	44
		154 (B) ^c	296
α -cumyl hydroperoxide	[80-15-9]	158 (B) ^c	296
<i>tert</i> -amyl hydroperoxide	[3425-61-4]	165 (B) ^c	296
<i>tert</i> -butyl hydroperoxide	[75-91-2]	172 (B) ^c	296
methyl ethyl ketone peroxide solutions	[1338-23-4]	d	292
2,4-pentanedione peroxide solutions	[37187-22-7]	d	292
peracetic acid	[79-21-0]	e	34,165
3-chloroperoxybenzoic acid	[937-14-4]	e	18

^a General discussions of decomposition temperatures of organic peroxides are given in Refs. 14, 21, 22, and 44.

^b 0.2 *M* solution in the solvent indicated in parentheses: AC = acetone, AMS = α -methylstyrene, B = benzene, D = dodecane, DC = decane, EA = ethyl acetate, OMS = odorless mineral spirits, and TCE = trichloroethylene.

^c Hydroperoxides are generally used with reducing agents, eg, iron salts, in redox emulsion polymerization systems.

^d These peroxides are used with activators, eg, cobalt carboxylates, and half-life data are of little significance.

^e Used as an epoxidizing, oxidizing, or bleaching agent; half-life data are of little significance.

Excluding the peroxyacids, which are used primarily as epoxidizing and bleaching agents, approximately 90% of the commercial organic peroxides are consumed by the polymer industry. The 1993 consumption of organic peroxides in the United States and in the world according to peroxide type is given in Table 18. In 1993, more than 103,000 metric tons of organic peroxides were consumed in the world, more than 33,000 metric tons in the United States. The three primary producers of organic peroxides in the United States are Elf Atochem North America, Inc., AKZO, and Laporte Industries. In the world the leading producers are AKZO, Elf Atochem North America, Inc., and Laporte Industries. Other producers include Witco Chemical Corp., Hercules, Inc., Norac Co., Reichhold Chemicals, Inc., Freeman Resins, ARCO Chemical Co., Nippon Oil & Fats Co., and Sanken Kako.

Table 18. Estimated 1993 Consumption of Organic Peroxides by Principal Types, ^a t

Peroxide type	United States	Worldwide
diacyl peroxides	6,400	19,000
peroxydicarbonates	2,900	8,000
hydroperoxides ^b	1,700	5,500
dialkyl peroxides	4,100	13,000
methyl ethyl ketone peroxides	7,000	22,000
peroxyesters	9,300	29,000
peroxyketals and others	2,300	7,100
<i>Total</i>	<i>33,700</i>	<i>103,600</i>

^a Estimates from Elf Atochem North America, Inc.
^b Does not include hydroperoxides produced for captive use.

Organic peroxides are used in the polymer industry as thermal sources of free radicals. They are used primarily to initiate the polymerization and copolymerization of vinyl and diene monomers, eg, ethylene, vinyl chloride, styrene, acrylic acid and esters, methacrylic acid and esters, vinyl acetate, acrylonitrile, and butadiene (see INITIATORS). They are also used to cure or cross-link resins, eg, unsaturated polyester–styrene blends, thermoplastics such as polyethylene, elastomers such as ethylene–propylene copolymers and terpolymers and ethylene–vinyl acetate copolymer, and rubbers such as silicone rubber and styrene–butadiene rubber.

The decomposition kinetics of an organic peroxide, as judged by 10-h HLT, largely determines the suitability of a particular peroxide initiator in an end use application (22). Other important factors are melting point, solubility, cost, safety, efficiency, necessity for refrigerated storage and shipment, compatibility with production systems, effects on the finished product, and potential for activation.

Free-radical reactions are accomplished using a variety of processes with different temperature requirements, eg, vinyl monomer polymerization and polymer modifications such as curing, cross-linking, and vis-breaking. Thus, the polymer industries are offered many different, commercial, organic peroxides representing a broad range of decomposition temperatures, as shown in Table 17 (19,22,31).

Organic peroxides also are used as flame-retardant synergists for polystyrene (306), for preparing block and graft copolymers, for reactive processing, for reducing the molecular weight of polypropylene (ie, controlled rheology or vis-breaking), for curing adhesives, for drying alkyd resin films, and for initiating cationic polymerization with cyclic ethers and maleic anhydride (44). Di- and triperoxides, which contain at least two peroxide moieties having different 10-h HLTs, decompose sequentially in the presence of vinyl monomers, yielding peroxy polymers and block copolymers (307,308). Organic peroxide initiators containing attached uv light absorbers, hindered amine light stabilizers (HALS), and antioxidants yield polymers to which uv-light stabilizer, HALS, and antioxidant groups are bound (309–311). Organic peroxides containing allyl groups have been used as molecular weight regulators in vinyl monomer polymerizations (312) and for attachment of epoxide groups to polymers and copolymers (313–315).

BPO is the preferred bleaching agent for flour and has been used to bleach gums, waxes, fats, and oils. It is the active ingredient in many acne medications. Diacyl peroxides have been used as burnout agents for acetate yarn, drying agents for Chinawood oils, and as free-radical sources in many organic syntheses (44). Di-*tert*-butyl peroxide is used as an ignition accelerator for diesel fuels and has been used in many organic syntheses either as a source of *tert*-butoxy (photo) or methyl (thermal) radicals.

BIBLIOGRAPHY

"Peroxides and Peroxy Compounds, Organic" in *ECT* 1st ed., Vol. 10, pp. 58–88, by N. A. Milas, Massachusetts Institute of Technology, P. H. Margulies and E. H. Stanley, Buffalo Electro-Chemical Co., Inc.; in *ECT* 2nd ed., Vol. 14, pp. 766–820, by O. L. Mageli and C. S. Sheppard, Wallace and Tiernan, Inc.; in *ECT* 3rd ed., Vol. 17, pp. 27–90, by C. S. Sheppard and O. L. Mageli, Pennwalt Corp.

1. D. Swern, ed., *Organic Peroxides*, Vol. I, Wiley-Interscience, New York, 1970.
2. D. Swern, ed., *Organic Peroxides*, Vol. II, Wiley-Interscience, New York, 1971.
3. D. Swern, ed., *Organic Peroxides*, Vol. III, Wiley-Interscience, New York, 1972.
4. A. G. Davies, *Organic Peroxides*, Butterworths, London, 1961.
5. E. G. E. Hawkins, *Organic Peroxides*, E. and F. F. Spon Ltd., London, 1961.
6. A. V. Tobolsky and R. B. Mesrobian, *Organic Peroxides*, Interscience Publishers, New York, 1954.
7. W. Adam, in W. Ando, ed., *Organic Peroxides*, John Wiley & Sons, Inc., New York, 1992, p. XV; A. J. Bloodworth and A. Shah, *Tetrahedron Lett.* **34**, 6643 (1993).
8. E. Okuyama, K. Umeyama, Y. Saito, M. Yamazaki, and M. Satake, *Chem. Pharm. Bull.* **41**, 1309 (1993).
9. S. W. Benson and R. Shaw, in Ref. 1, Chapt. II, p. 128.
10. R. Hiatt, in Ref. 2, Chapt. I, pp. 45–46.
11. J. C. Masson, in J. Brandrup and E. H. Immergut, eds., *Polymer Handbook*, 3rd ed., John Wiley & Sons, Inc., New York, 1989, pp. 1–65.
12. D. Griller and K. U. Ingold, *Acc. Chem. Res.* **9**, 13 (1976).
13. C. J. Dyball, R. B. Gallagher, V. R. Kamath, and S. E. Stromberg, *Plast. Des. Proc.*, 41 (Jan. 1980).
14. C. S. Sheppard, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*. 2nd ed., Vol. 11, Wiley-Interscience, New York, 1988, pp. 1–21.
15. P. A. Callais, in D. Satas, ed., *Coatings Technology Handbook*, Marcel Dekker, Inc., New York, 1991, Chapt. 60, pp. 522–527.
16. G. A. Harpell and D. H. Walrod, *Rubber Chem. Technol.* **46**, 1007 (1973).
17. G. A. Razuvaev and T. G. Brilkina, *Russ. Chem. Rev.* **45**, 1135 (1976).
18. M. Yamada, K. Kitagawa, and T. Komai, *Plast. Ind. News* **17**, 131 (1971).
19. R. L. Friedman, R. N. Lewis, and R. L. Pastorino, *Mod. Plast.*, 66 (Nov. 1971).
20. F. C. Baines, J. H. Grezlak, and A. V. Tobolsky, *J. Polym. Sci. A-1* **7**, 3297 (1969).
21. C. S. Sheppard and V. R. Kamath, *Polym. Eng. Sci.* **19**, 597 (1979).
22. *Technical Publication, HALFLIFE—Peroxide Selection Based on Half-Life*, 2nd ed., and Copyrighted Software, Elf Atochem North America, Inc., Philadelphia, Pa., 1992; P. A. Callais, V. R. Kamath, S. E. Stromberg, J. Sanchez, and J. M. Gravelle, *Plast. Compound.* **15**(1), 49–52 (1992).
23. *Free Radical Initiators for the Suspension Polymerization of Vinyl Chloride*, Technical data, Elf Atochem North America, Inc., Philadelphia, Pa., Sept. 1993.
24. J. D. van Drumpt, *Technical Publication—Crosslinking of Polyolefins: A General Introduction*, Akzo Chemicals Inc., Chicago, Ill., 1986.
25. "Organic Peroxides: Their Safe Handling and Use," Technical Bulletin 30.40, *Organic Peroxides*, Elf Atochem North America, Inc., Philadelphia, Pa., Mar. 1994.
26. "Safe Handling, Storage and Transportation of Peroxides Requiring Refrigeration," Technical Bulletin 30.43, *Organic Peroxides*, Elf Atochem North America, Inc., Philadelphia, Pa., Apr. 1993.
27. W. A. Pryor, ed., *Organic Free Radicals*, ACS Symposium Series 69, American Chemical Society, Washington, D. C., 1978.
28. N. A. Sokolov and Yu. A. Aleksandrov, *Russ. Chem. Rev.* **47**, 172 (1978).
29. K. Matsukura and M. Yamaguchi, *Nippon Gomu Kyokaishi* **50**, 625 (1977).
30. W. M. Weigert and P. Kleinschmit, in W. M. Wiegert, ed., *Wasserstoffperoxide Seine Derivate*, Heuthig, Heidelberg, Germany, 1978, Chapt. III, pp. 37–72.
31. G. Luft and H. Bitsch, *J. Macromol. Sci. Chem.* **A11**, 1089 (1977); G. Luft, P. Mehrling, and H. Seidl, *Angew. Makromol. Chem.* **73**, 95 (1978); G. Luft, H. Grimm, and M. Dorn, *Die Angew. Makromol. Chem.* **174**, 165 (1990).
32. O. L. Mageli and C. S. Sheppard, in Ref. 1, Chapt. I refs. 17, 19, 48, 50–54, 56–64, 66–103, 190, 196, and 257.
33. G. Sosnovsky and J. H. Brown, *Chem. Rev.* **66**, 529 (1966).
34. J. O. Edwards, ed., *Peroxide Reaction Mechanisms*, John Wiley & Sons, Inc., New York, 1962.
35. W. Ando, ed., *Organic Peroxides*, John Wiley & Sons, Inc., New York, 1992.
36. S. Patai, ed., *The Chemistry of Peroxides*, John Wiley & Sons, Inc., New York, 1983; S. Patai, ed., *Supplement E2: The Chemistry of Hydroxyl, Ether and Peroxide Groups*, John Wiley & Sons, New York, 1993.
37. H. Kropf, ed., *Organische Peroxo-Verbindungen*, Georg Thieme, New York, 1988.

38. A. L. Baumstark, ed., *Advances in Oxygenated Processes*, Vol. 1, JAI Press, Greenwich, Conn., 1988.
39. W. Ando and Y. Moro-oka, eds., *The Role of Oxygen in Chemistry and Biochemistry*, Elsevier, New York, 1988.
40. W. A. Herrmann, ed., *Organic Peroxygen Chemistry, Topics in Current Chemistry 164*, Springer-Verlag, Berlin, 1993.
41. L. F. Martin, *Organic Peroxide Technology*, Noyes Data Corp., Park Ridge, N.J., 1973.
42. B. C. Brodie, *Ann.* **108**, 79 (1858).
43. Yu. A. Alexandrov, *J. Organometallic Chem.* **238**, 1 (1982).
44. O. L. Mageli and C. S. Sheppard, in Ref. 1, Chapt. I, pp. 1–104.
45. N. A. Porter, in Ref. 35, Chapt. 2, pp. 101–156.
46. L. S. Silbert, in Ref. 2, Chapt. VII, pp. 637–798.
47. G. A. Razuvaev, V. A. Shushunov, V. A. Dodonov, and T. G. Brilkina, in Ref. 3, Chapt. III, pp. 141–270.
48. G. Sosnovsky and D. J. Rawlinson, in Ref. 2, Chapt. II, pp. 153–268.
49. E. Huf, in Ref. 40, pp. 69–76; K. B. Sharpless and T. R. Verhoeven, *Aldrichemica Acta* **12**(4), 63 (1979).
50. E. Rizzardo and D. H. Solomon, *J. Macromol. Sci. Chem.* **A-14**, 33 (1980).
51. J. K. Kochi, ed., *Free Radicals*, Vol. I, John Wiley & Sons, Inc., New York, 1973.
52. C. Walling, *Free Radicals in Solution*, John Wiley & Sons, Inc., New York, 1957.
53. J. A. Howard, in Ref. 27, Chapt. 25, pp. 413–432.
54. E. S. Shanley and F. P. Greenspan, *Ind. Eng. Chem.* **39**, 1536 (1947); T. A. Schenach, *Chem. Eng. News*, 39 (Feb. 5, 1973); W. Edl, *Chem. Eng. News*, 40, 49 (May 14, 1973).
55. J. K. Kochi, ed., *Free Radicals*, Vol. II, John Wiley & Sons, Inc., New York, 1973.
56. N. R. Subbaratham and J. G. Calvert, *J. Am. Chem. Soc.* **84**, 1113 (1962); S. S. Thomas and J. G. Calvert, *J. Am. Chem. Soc.* **84**, 4207 (1962).
57. M. L. Kaplan, *Chem. Tech.*, 621 (Oct. 1971).
58. M. Stratakis and M. Orfanopoulos, *Synthetic Comm.* **23**, 425 (1993).
59. **U.S. Pat. 4,010,152** (Mar. 1, 1977), R. E. MacLeay and C. S. Sheppard (to Pennwalt Corp.).
60. Yu. A. Aleksandrov, S. A. Lebedev, and V. A. Alferov, *Zh. Obshch. Khim.* **48**, 1660 (1978).
61. W. Adam and J. del Fierro, *J. Org. Chem.* **43**, 1159 (1978).
62. J. Rigaudy and G. Izoret, *Compt. Rend.* **236**, 2086 (1953).
63. B. Plesnicar, in Ref. 35, Chapt. 10, pp. 479–533.
64. K. J. McCullough and M. Nojima, in Ref. 35, Chapt. 13, pp. 661–728.
65. W. A. Pryor, N. Ohto, and D. F. Church, *J. Am. Chem. Soc.* **104**, 5813 (1982).
66. R. Hiatt, in Ref. 3, Chapt. I, pp. 1–65.
67. P. Rademacher and W. Elling, *Lieb. Ann. Chem.*, 1473 (1979).
68. R. Criegee and G. Paulig, *Chem. Ber.* **88**, 712 (1955).
69. S. Matsugo, in Ref. 35, Chapt. 3, pp. 157–194.
70. W. H. Richardson, R. McGinness, and H. E. O'Neal, *J. Org. Chem.* **46**, 1887 (1981).
71. W. Adam and J. Sanabia, *J. Am. Chem. Soc.* **99**, 2735 (1977).
72. Technical data, Elf Atochem North America, Inc., Philadelphia, Pa.
73. F. Tang and E. S. Huyser, *J. Org. Chem.* **43**, 1016 (1978).
74. R. Hiatt and P. M. Rahimi, *Int. J. Chem. Kinet.* **10**(3), 185 (1978).
75. S. K. Wong, *J. Am. Chem. Soc.* **101**, 1235 (1979).
76. Yu. N. Baryshnikov, G. I. Vesnovskysya, L. N. Kibkalo, and V. N. Koshlakova, *J. Org. Chem. USSR* **15**, 838 (1979).
77. W. H. Richardson and co-workers, *J. Org. Chem.* **38**, 4219 (1973).
78. L. M. Mudraya, A. I. Rakhimov, F. V. Mudryi, and A. I. Lemasov, *Zh. Obshch. Khim.* **49**, 477 (1979).
79. K. Kishore, V. Gayathri, and K. Ravindran, *J. Macromol. Sci. Chem.* **A16**(8) 1359 (1981).
80. A. A. Frimer, *Chem. Rev.* **79**, 359 (1979).
81. E. L. Clennan and C. S. Foote, in Ref. 35, Chapt. 6, pp. 253–318.
82. M. Suzuki, H. Ohtake, Y. Kameya, N. Hamanaka, and R. Noyori, *J. Org. Chem.* **54**, 5292 (1989).
83. M. Balci, *Chem. Rev.* **81**, 91 (1981).
84. M. Natsume, Y. Sekine, and M. Ogawa, *Tetrahedron Lett.* (36), 3473 (1979); M. Natsume and H. Muratake, *Tetrahedron Lett.* (36), 3477 (1979).
85. A. Adam, M. Heil, T. Mosandt, and C. R. Saha-Müller, in Ref. 35, Chapt. 5, pp. 221–254.
86. W. Adam and A. Treiber, *J. Org. Chem.* **59**, 840 (1994).
87. L. A. Hull and L. Budhai, *Tetrahedron Lett.* **34**, 5039 (1993).
88. A. Adam, L. P. Hadjarapoglou, R. Curci, and R. Mello, in Ref. 35, Chapt. 4, pp. 195–219.
89. R. W. Murray, *Chem. Rev.* **89**, 1187 (1989).
90. W. Adam, R. Curci, and J. O. Edwards, *Acc. Chem. Res.* **22**, 205 (1989).
91. R. Mello and co-workers, *J. Am. Chem. Soc.* **113**, 2205 (1991).
92. M. Singh and R. W. Murray, *J. Org. Chem.* **57**, 4263 (1992).
93. C. W. Jefford and C. G. Rimbault, *J. Am. Chem. Soc.* **100**, 6437 (1978).
94. V. A. Yablokov, *Russian Chem. Rev.* **49**, 833 (1980).
95. S. Matsubara, K. Takai, and H. Nozaki, *Bull. Chem. Soc. Jpn.* **56**, 2029 (1983).
96. S. Florio and L. Troisi, *Tetrahedron Lett.* **34**, 3141 (1993).
97. **U.S. Pat. 4,133,835** (Jan. 9, 1979), R. A. Bafford and O. L. Mageli (to Pennwalt Corp.).
98. "Dialkyl Peroxides," Product Bulletin, *Organic Peroxides*, Elf Atochem North America, Inc., Philadelphia, Pa., Jan. 1993.
99. *Technical Bulletin ORC-301K*, Hercules Inc., Wilmington, Del.
100. A. J. Bloodworth and R. J. Bunce, *J. Chem. Soc. Perkin I*, 2787 (1972).
101. R. A. Johnson, E. G. Nidy, and M. V. Merritt, *J. Am. Chem. Soc.* **100**, 7960 (1978); C. I. Chern, R. DiCosimo, R. De Jesus, and J. San Filippo, *J. Am. Chem. Soc.* **100**, 7317 (1978).
102. I. B. Afanas'ev, *Russ. Chem. Rev.* **48**, 527 (1979).
103. M. V. George and V. Bhat, *Chem. Rev.* **79**, 447 (1979).
104. K. A. Horn and G. B. Schuster, *J. Am. Chem. Soc.* **100**, 6649 (1978).
105. **U.S. Pat. 3,632,606** (Jan. 4, 1972), R. L. Talbott and P. G. Thompson (to Minnesota Mining and Manufacturing Co.).
106. J. O. Edwards, R. H. Pater, R. Curci, and F. Di Furia, *Photochem. Photobiol.* **30**, 63 (1979).
107. R. W. Murray and R. Jeyaraman, *J. Org. Chem.* **50**, 2847 (1985); W. Adam and L. Hadjarapoglou, in Ref. 40, pp. 45–62.
108. M. F. Salomon and R. G. Salomon, *J. Am. Chem. Soc.* **101**, 4290 (1979).
109. T. Mukundan and K. Kishore, *Prog. Polym. Sci.* **15**, 475 (1990).
110. **U.S. Pat. 5,039,754** (Aug. 13, 1991), J. Sanchez (to Atochem North America, Inc.).
111. B. M. Monroe, *J. Phys. Chem.* **82**, 15 (1978).
112. J. R. Sanderson, A. G. Zeiler, and R. J. Wilterdink, *J. Org. Chem.* **40**, 2239 (1975).

113. N. A. Porter, J. R. Nixon, and D. W. Gilmore, in Ref. 27, Chapt. 6, pp. 89–101.
114. W. Adam and H. J. Eggelte, *Angew. Chem.* **89**, 762 (1977).
115. A. G. Davies, in Ref. 2, Chapt. IV, pp. 337–354.
116. M. S. Kharasch and G. Sosnovsky, *J. Org. Chem.* **23**, 1322 (1958).
117. J. R. Sanderson, R. J. Wilterdink, and A. G. Zeiler, *Synthesis*, 479 (1976).
118. P. R. Story and P. Busch, *Adv. Org. Chem.* **8**, 1 (1972).
119. P. Busch and P. R. Story, *Synthesis*, 181 (1970).
120. M. J. C. Harding and D. M. Whalen, *Ind. Eng. Chem. Prod. Res. Dev.* **14**, 232 (1975).
121. P. R. Story, B. Lee, C. E. Bishop, D. D. Denson, and P. Busch, *J. Org. Chem.* **35**, 3059 (1970).
122. M. Schulz and K. Kirschke, in Ref. 3, Chapt. II, pp. 67–140.
123. C. T. Ratcliffe, C. V. Hardin, L. R. Anderson, and W. B. Fox, *J. Chem. Soc., Chem. Commun.*, 784 (1971).
124. E. G. E. Hawkins, *J. Chem. Soc.*, 3463 (1955).
125. P. S. Bailey, *Chem. Rev.* **58**, 925 (1958).
126. A. T. Menyailo and M. V. Pospelov, *Russ. Chem. Rev.* **36**, 284 (1967).
127. W. P. Keaveney, M. G. Berger, and J. J. Pappas, *J. Org. Chem.* **32**, 1537 (1967).
128. P. Dussault, *Chem. Eng. News* **71**, 2 (1993).
129. R. W. Murray and S. K. Agarwal, *J. Org. Chem.* **50**, 4698 (1985).
130. R. Criegee and G. Wenner, *Ann.* **564**, 9 (1949).
131. P. Dussault and A. Sahli, *J. Org. Chem.* **57**, 1009 (1992).
132. I. Saito, R. Nagata, K. Yuba, and T. Matsuura, *Tetrahedron Lett.* **24**, 1737 (1983).
133. V. J. Karnojitzky, *Chim. Ind. (Paris)* **88**, 233 (1962).
134. A. H. Owen-Jones, *Firemen* **33**, 15 (1975).
135. Th. M. Wortel and H. van Bakkum, *J. Org. Chem.* **45**, 4763 (1980).
136. A. I. Rakhimov, V. V. Chapurkin, V. I. Filyakova, K. I. Pashkevich, and S. L. Kokorin, *Zh. Org. Khim.* **27**, 409 (1991).
137. P. S. Starcher, B. Phillips, and F. C. Frostick, *J. Org. Chem.* **26**, 3568 (1961).
138. **Ger. Pat. 1,100,033** (Feb. 23, 1961), A. Rieche, E. Schmitz, and E. Gröndemann.
139. **U.S. Pat. 3,576,826** (Apr. 27, 1971), R. A. Bafford, L. E. Korczykowski, and O. L. Mageli (to Pennwalt Corp.).
140. C. W. Jefford and co-workers, *Stud. Org. Chem. (Amsterdam)* **31**, 113 (1987); C. W. Jefford, *Stud. Surf. Sci. Catal.* **66**, 555 (1991).
141. C. W. Jefford, J. C. Rossier, and G. D. Richardson, *J. Chem. Soc., Chem. Commun.*, 1064 (1983).
142. J. Sanchez, *ACS 9th Middle Atlantic Regional Meeting*, Wilkes-Barre, Pa., Paper 108, Apr. 1974.
143. C. W. Jefford and co-workers, *Synth. Comm.* **20**, 2589 (1990).
144. Canadian Pat. 730,181 (March 15, 1965), N. A. Milas, (to Massachusetts Institute of Technology).
145. **U.S. Pat. 4,217,290** (Aug. 12, 1980), P. Maggioni, F. Minisci, and A. Citterio (to Brichima SpA).
146. Y. Sugihara and co-workers, *Bull. Chem. Soc. Jpn.* **65**, 664 (1992); Y. Watanabe, K. Ohta, and S. Suyama, *Bull. Chem. Soc. Jpn.* **65**, 2063 (1992); Y. Watanabe, H. Ishigaki, and S. Suyama, *Polymer J.* **24**(3), 257 (1992); Y. Watanabe, Y. Sugihara, S. Suyama, and T. Nakamura, *Polymer J.* **24**(9), 971 (1992); K. Matsuyama and H. Kumura, *J. Org. Chem.* **58**, 1766 (1993).
147. R. E. Drumright, P. E. Kastl, and D. B. Priddy, *Macromolecules* **26**, 2246 (1993); R. E. Drumright, E. Ellington, P. E. Kastl, and D. B. Priddy, *Macromolecules* **26**, 2253 (1993); V. A. Dias, R. E. Drumright, E. Ellington, P. E. Kastl, and D. B. Priddy, *Macromolecules* **26**, 2259 (1993).
148. **Jpn. Pat. 93/194621** (1993), (to Nippon Oils and Fats Co., Ltd.).
149. **U.S. Pat. 3,149,126** (Sept. 15, 1964), N. A. Milas (to Massachusetts Institute of Technology); **U.S. Pat. 3,349,040** (Oct. 24, 1967), D. M. Kressin and co-workers (to Wallace & Tiernan Inc.); **U.S. Pat. 3,377,407** (Apr. 9, 1968), D. M. Kressin and co-workers (to Wallace & Tiernan Inc.); N. A. Milas and co-workers, *J. Am. Chem. Soc.* **85**, 222 (1963).
150. **U.S. Pat. 2,601,224** (June 24, 1952), M. J. Roedel (to E.I. du Pont de Nemours & Co., Inc.).
151. A. Noponen, *Chem. Eng. News*, 5 (Feb. 21, 1977).
152. J. R. Sanderson and P. R. Story, *J. Org. Chem.* **39**, 3463 (1974).
153. J. R. Sanderson, P. R. Story, and P. Kalidas, *J. Org. Chem.* **40**, 691 (1975).
154. P. Rademacher and W. Ellign, *Lieb. Ann. Chem.*, 1473 (1979).
155. F. A. L. Anet and I. Yavari, *Tetrahedron Lett.* **42**, 3787 (1976).
156. **Ger. Pat. 1,518,128** (Oct. 22, 1970), C. Röchardt (to Badische Anilin & Soda Fabrik AG).
157. C. Röchardt and S. Rochlitz, *Ann.*, 15 (1974).
158. C. Röchardt, G. Hamprecht, and H. Brinkmann, *Chem. Ber.* **108**, 3210 (1975).
159. **U.S. Pat. 3,983,131** (Sept. 28, 1976), J. Trüger, O. Weiberg, and W. Weigert (to Deutsche Gold-und Silber-Scheideanstalt).
160. S.-I. Murahashi and co-workers, *J. Am. Chem. Soc.* **112**, 7820 (1990).
161. R. Criegee, *Chimia* **22**, 392 (1968).
162. P. S. Bailey and T. M. Ferrell, *J. Am. Chem. Soc.* **100**, 899 (1978).
163. W. Adam, H. J. Eggelte, and A. Rodriguez, *Synthesis*, 383 (1979).
164. R. D. Suenram and F. J. Lovas, *J. Am. Chem. Soc.* **100**, 5117 (1978).
165. D. Swern, in Ref. 1, Chapt. VI, pp. 313–474; D. Swern, in Ref. 1, Chapt. VII, pp. 475–516; D. Swern, in Ref. 2, Chapt. V, pp. 355–533.
166. A. F. Hegarty, *Comp. Org. Chem.* **2**, 1105 (1979).
167. K. Sasaki, *Sekiyu Gakkai Shi* **19**, 421 (1977); S. N. Lewis in R. L. Augustine, ed., *Oxidation*, Vol. 1, Marcel Dekker, Inc., New York, 1969, Chapt. 5, pp. 213–258.
168. Y. Sawaki, in Ref. 35, Chapt. 9, pp. 425–477.
169. E. Koubek and J. E. Welsch, *J. Org. Chem.* **33**, 445 (1968).
170. R. Curci and J. O. Edwards, in Ref. 1, Chapt. IV, pp. 199–264.
171. M. Botsivali and D. F. Evans, *J. Chem. Soc. Chem. Comm.*, 1114 (1979).
172. D. F. Evans and M. W. Upton, *J. Chem. Soc., Dalton Trans.* **6**, 1151–1158 (1985).
173. **U.S. Pat. 4,957,647** (Sept. 18, 1990), A. G. Zielske (to The Clorox Co.).
174. G. Ya. Perchugov, N. A. Sokolov, and Yu. A. Aleksandrov, *Dokl. Akad. Nauk SSSR* **235**(1), 121 (1977).
175. J. Fossey, D. Lefort, and J. Sorba, in Ref. 40, pp. 99–113.
176. Y. Ogata and K. Tomizawa, *J. Org. Chem.* **43**, 261, (1978); C. F. Hendriks, H. C. A. van Beek, and P. M. Heertjes, *Ind. Eng. Chem. Prod. Res. Dev.* **18**, 38 (1979).
177. H. Muecke, *Pharmazie* **32**, 612 (1977).
178. G. O. Bagdasaryan, E. G. Garibyan, I. A. Vardanyan, and A. B. Nalbandyan, *Arm. Khim. Zh.* **30**, 447 (1977).
179. I. J. Harvie and F. J. McQuillin, *J. Chem. Soc. Comm.*, 369 (1976).
180. G. O. Bagdasaryan, I. A. Vardanyan, and A. B. Nalbandyan, *Dokl. Akad. Nauk SSSR* **231**, 362 (1976).
181. H. Heaney, in Ref. 40, pp. 1–19; H. Heaney, *Aldrichemica Acta* **26**(2), 35 (1993).
182. **Brit. Pat. 1,550,165** (Aug. 8, 1979), O. Weiberg, R. Wirthwein, H. Seifert, and G. Siekmann (to Deutsche Gold-und Silber-Scheideanstalt Vormals

- Roessler and Bayer Aktiengesellschaft); **U.S. Pat. 4,089,892** (May 16, 1978), W. Hofen, G. Schreyer, R. Wirthwein, H. Waldmann, and G. Siekmann (to Deutsche Gold-und Silber-Scheideanstalt Vormals Roessler and Bayer Aktiengesellschaft).
183. S. A. Maslov and E. A. Blyumberg, *Russ. Chem. Rev.* **45**, 155 (1976).
 184. J. Y. Nedelec, J. Sorba, and D. Lefort, *Synthesis* **12**, 821 (1976).
 185. R. M. Coates and J. M. Williams, *J. Org. Chem.* **39**, 3054 (1974).
 186. P. A. Grieco, Y. Yokoyama, S. Gilman, and M. Nishizawa, *J. Org. Chem.* **42**, 2034 (1977).
 187. R. Hiatt, in Ref. 2, Chapt. VIII, pp. 799–929.
 188. K. Fujimori, in Ref. 35, Chapt. 7, pp. 319–385.
 189. **U.S. Pat. 5,034,485** (July 23, 1991), D. B. Priddy and W. H. Tong (to The Dow Chemical Company); **U.S. Pat. 5,087,738** (Feb. 11, 1992), D. B. Priddy and W. H. Tong (to The Dow Chemical Company).
 190. **U.S. Pats. 3,799,966** (Mar. 26, 1974), 3,855,351 (Dec. 17, 1974), and 3,857,828 (Dec. 31, 1974), J. Sanchez (to Pennwalt Corp.).
 191. **U.S. Pat. 3,657,311** (Apr. 18, 1972), A. J. D'Angelo (to Pennwalt Corp.).
 192. **U.S. Pat. 3,657,312** (Apr. 18, 1972), A. J. D'Angelo (to Pennwalt Corp.).
 193. **U.S. Pat. 3,499,919** (Mar. 10, 1970), H. G. Gerritsen, H. Hansma, and H. Jaspers (to Noury & van der Lande NV).
 194. **U.S. Pat. 3,720,700** (Mar. 13, 1973), H. Norback (to Kema Nord AB, Sweden).
 195. **U.S. Pat. 3,586,722** (June 22, 1971), J. Sanchez (to Pennwalt Corp.).
 196. **U.S. Pat. 4,043,940** (Aug. 23, 1977), J. Sanchez (to Pennwalt Corp.).
 197. **U.S. Pat. 3,650,972** (Mar. 21, 1972), J. Sanchez (to Pennwalt Corp.).
 198. J. D. Coyle, *Chem. Rev.* **78**, 97 (1978).
 199. R. Hiatt, *Intra Sci. Chem. Rep.* **5**, 163 (1971).
 200. J. E. Lefler and A. A. More, *J. Am. Chem. Soc.* **94**, 2483 (1972).
 201. H. Sawada, *Rev. Heteroat. Chem.* **8**, 205 (1993).
 202. J. A. Barter and D. E. Kellar, *J. Polym. Sci. Polym. Chem. Ed.* **15**, 2545 (1977).
 203. G. S. Bylina, M. S. Matveentseva, and Yu. A. Ol'dekop, *Vysokomol. Soedin Ser. B* **19**, 555 (1977).
 204. **U.S. Pat. 4,079,074** (Mar. 14, 1978), J. Sanchez, V. R. Kamath, and J. C. Halas (to Pennwalt Corp.).
 205. P. B. Dervan and C. R. Jones, *J. Org. Chem.* **44**, 2116 (1979).
 206. T. A. Nadervel, V. P. Sass, S. I. Serov, and S. V. Sokolov, *J. Org. Chem. USSR* **13**, 2136 (1977); V. P. Sass, S. I. Serov, and S. V. Sokolov, *J. Org. Chem. USSR* **13**, 2140 (1977).
 207. W. C. Danen and R. L. Arudi, *J. Am. Chem. Soc.* **100**, 3944 (1978).
 208. *Peroxydicarbonates*, Product Bulletin, Organic Peroxide Division, Atochem North America, Inc., Buffalo, N.Y., Nov. 1991.
 209. *Organosulfonyl Peroxides*, Product Bulletin, Lucidol Division, Pennwalt Corp., Buffalo, N.Y., 1977.
 210. **U.S. Pat. 3,580,955** (May 25, 1971), R. A. Bafford (to Pennwalt Corp.).
 211. T. Suehiro, Y. Matsui, and Y. Inoguchi, *Bull. Chem. Soc. Jpn.* **51**, 2609 (1978).
 212. **U.S. Pat. 4,069,239** (Jan. 17, 1978), R. N. Lewis, R. L. Pastorino, and J. F. Wilts (to Argus Chemical Corp.).
 213. I. A. Singer, in Ref. 1, Chapt. V, pp. 265–312.
 214. W. Adam, A. Alzerreca, Ju-C. Liu, and F. Yany, *J. Am. Chem. Soc.* **99**, 5768 (1977).
 215. A. I. Rakhimov, O. S. Bogdanova, and A. A. Okolelova, *Zh. Org. Khim.* **14**, 1338 (1978).
 216. **USSR Pat. 520,352** (July 5, 1976), M. Kh. Myasoedova, T. F. Rakhimova, and A. I. Rakhimov.
 217. **U.S. Pat. 5,314,970** (May 24, 1994), R. T. Kazmierczak, J. P. Hibbard, and R. E. MacLeay (to Elf Atochem North America, Inc.).
 218. **U.S. Pat. 4,119,657** (Oct. 10, 1978), T. Komai and K. Matsuyama (to Nippon Oils & Fats Co., Ltd.).
 219. **U.S. Pat. 4,954,656** (Sept. 4, 1990), S. Suyama and Y. Higuchi (to Nippon Oils & Fats Co., Ltd.); **U.S. Pat. 5,041,624** (Aug. 20, 1991), S. Suyama, H. Ishigaki, and K. Taura (to Nippon Oils & Fats Co., Ltd.); **U.S. Pat. 5,258,465** (Nov. 2, 1993), S. Suyama, Y. Watanabe, H. Ishigaki, and H. Kumura (to Nippon Oils & Fats Co., Ltd.).
 220. **U.S. Pat. 5,292,914** (Mar. 8, 1994), S. Suyama, T. Nakamura, H. Nagai, and M. Takei (to NOF Corp.).
 221. N. Uyanik and B. M. Baysal, *J. Appl. Polym. Sci.* **41**, 1981 (1990).
 222. P. D. Bartlett, E. P. Benzing, and R. E. Pincock, *J. Am. Chem. Soc.* **82**, 1762 (1960); P. D. Bartlett and R. E. Pincock, *J. Am. Chem. Soc.* **82**, 1769 (1960).
 223. **U.S. Pat. 4,859,794** (Aug. 22, 1989), C. O. A. Lundin, E. M. Larsson, and B. J. E. Akermark (to Berol Nobel Nacka AB); *KENODOX, Polymerization Initiator for Vinyl Chloride*, Experimental Product Bulletin, Berol Nobel, Stockholm, Sweden.
 224. **U.S. Pat. 4,525,308** (June 25, 1985), J. Sanchez (to Pennwalt Corp.); **U.S. Pat. 4,634,753** (Jan. 6, 1987), J. Sanchez (to Pennwalt Corp.).
 225. **U.S. Pat. 3,435,060** (Mar. 25, 1969), C. J. Rychardt (to Elektrochemische Werke München AG).
 226. I. S. Krull and A. Mandelbaum, *Org. Mass. Spectrum* **11**, 504 (1976).
 227. N. A. Turovskii, A. A. Turovskii, and N. V. Dzumedzei, *Zh. Strukt. Khim.* **18**, 197 (1977).
 228. I. D. Zaikin, *Zh. Fiz. Khim.* **51**, 1330 (1977).
 229. J. J. Davis and J. G. Miller, *J. Am. Chem. Soc.* **99**, 245 (1977).
 230. L. V. Luk'yanenko, A. A. Turovskii, and N. A. Turovskii, *Zh. Obshch. Khim.* **49**, 1219 (1979).
 231. D. C. Neckers, *J. Radiat. Curing* **10**, 19 (1983).
 232. N. S. Allen and co-workers, *J. Appl. Polym. Sci.* **42**, 1169 (1991).
 233. P. D. Bartlett and C. Rychardt, *J. Am. Chem. Soc.* **82**, 1756 (1960).
 234. P. D. Bartlett and L. B. Gortler, *J. Am. Chem. Soc.* **85**, 1864 (1963).
 235. J. P. Lorand and P. D. Bartlett, *J. Am. Chem. Soc.* **88**, 3294 (1966).
 236. W. Duismann and C. Rychardt, *Lieb. Ann. Chem.*, 1834 (1976).
 237. J. C. Martin, in Ref. 27, Chapt. 5, pp. 71–88.
 238. W. Adam, O. Cueto, L. N. Guedes, and L. O. Rodriguez, *J. Org. Chem.* **43**, 1466 (1978).
 239. S. P. Schmidt and G. B. Schuster, *J. Am. Chem. Soc.* **100**, 1966 (1978).
 240. W. Adam and O. Cueto, *J. Am. Chem. Soc.* **101**, 6511 (1979).
 241. B. G. Dixon and G. B. Schuster, *J. Am. Chem. Soc.* **101**, 3116 (1979).
 242. T. Ochiai, Y. Usuda, M. Yoshida, and O. Simamura, *Bull. Chem. Soc. Jpn.* **49**, 2522 (1976).
 243. D. J. Rawlinson, M. Konieczny, and G. Sosnovsky, *Z. Naturforsch. B Anorg. Chem. Org. Chem.* **34B**, 76 (1979).
 244. C. Hepburn, *Rubber World* **190**, 49 (1984).
 245. A. C. M. Meesters and M. H. Benn, *Synthesis*, 679 (1978).
 246. **U. S. Pat. 4,131,728** (Dec. 26, 1978), D. B. Priddy (to Dow Chemical Company).
 247. **U.S. Pat. 3,950,375** (Apr. 13, 1976), P. M. McKee and U. D. Wagle (to Pennwalt Corp.).
 248. R. D. Mair and R. T. Hall, *Treatise Anal. Chem.* **14**, 295 (1971).
 249. A. J. Martin in J. Mitchell, Jr., ed., *Organic Analysis*, Vol. IV, Interscience Publishers Inc., New York, 1960, pp. 1–64.
 250. S. Siggia and J. G. Hanna, *Quantitative Organic Analysis via Functional Groups*, 4th ed., John Wiley & Sons, Inc., New York, 1979, pp. 325–372.
 251. R. M. Johnson and I. W. Siddiqi, *The Determination of Organic Peroxides*, Pergamon Press, New York, 1970.
 252. R. D. Mair and R. T. Hall, in Ref. 2, Chapt. VI, pp. 535–635.

253. ASTM Standard E 298-91, "Standard Test Method for Assay of Organic Peroxides," *1992 Annual Book of ASTM Standards*, Section 15, Vol. 15.05, ASTM, Philadelphia, Pa., 1992.
254. ASTM Standard E 1228-88, "Standard Test Method for Peroxy Esters-Catalyzed Iodometric Procedure," in Ref. 253.
255. L. Dulog and K. H. Burg, *Z. Anal. Chem.* **203**, 184 (1964).
256. R. M. Johnson and I. W. Siddiqi, *The Determination of Organic Peroxides*, Pergamon Press, New York, 1970.
257. T. I. Yurzhenko, A. A. Zan'ko, O. K. Serdyukova, and L. P. Mamchur, *Dokl. L'vovsk. Politekh. Inst.* **5**, 41 (1963).
258. V. N. Pavlov, N. A. Kurochkina, and Z. S. Frolova, *Zavodsk. Lab.* **30**, 539 (1964).
259. V. N. Dmitrieva, O. V. Meshkova, and V. D. Bezuglyi, *Zh. Analit. Khim.* **19**, 389 (1964).
260. Yu. N. Anisimov, S. S. Ivanchev, and A. I. Yurzhenko, *Zh. Analit. Khim.* **21**, 113 (1966).
261. S. W. Bukata, L. L. Zabrocki, and M. F. McLaughlin, *Anal. Chem.* **35**, 885 (1963).
262. M. P. Hahto and J. E. Beaulieu, *J. Chromatog.* **25**, 472 (1966).
263. ASTM Standard E 475-84, "Standard Test Method for Assay of Di-*tert*-Butyl Peroxide by Gas Chromatographic Analysis," in Ref. 253.
264. M. M. Buzlanova, V. F. Stepanovskaya, A. F. Nesterov, and V. L. Antonovskii, *Zh. Analit. Khim.* **21**, 506 (1966).
265. S. Hayano, T. Ota, and Y. Fukushima, *Bunseki Kagaku* **15**, 365 (1966).
266. G. Dobson and G. Hughes, *J. Chromatog.* **16**, 416 (1964).
267. W. Eggersgluss, *Organische Peroxyde*, Verlag Chemie, Weinheim, Germany, 1951.
268. ASTM Standard E 755-90, "Standard Test Method for Assay of Dicumyl Peroxide by Liquid Chromatography," in Ref. 253.
269. ASTM Standard E 299-90, "Standard Test Method for Trace Amounts of Peroxides in Organic Solvents," in Ref. 253.
270. ASTM Standard E 440-72, "Standard Test Method for Traces of Peroxides in Organic Solvents With N,N'-Diphenyl-*p*-Phenylenediamine," in Ref. 253.
271. E. P. Floyd and H. E. Stokinger, *AIHA J.* **19**, 205 (1958).
272. Von H. J. Kschle, *Zentralbl. Arbeitsmed. Arbeitsschutz* **8**, 25 (1958).
273. *SPI Bulletin 19-B*, 1992 rev., The Society of the Plastics Industry, Inc., New York and Washington, D.C.
274. J. B. Galvin and C. H. Farr, in G. D. Clayton and F. E. Clayton, eds., *Patty's Industrial Hygiene and Toxicology*, 4th ed., Vol. 2, Part A, John Wiley & Sons, Inc., 1993, Chapt. 9, pp. 527-597.
275. *t*-Butyl Perbenzoate, National Toxicology Program, Toxicity Report Series Number 15, NIH Publication 92-3134, United States Department of Health and Human Services, Washington, D.C., July 1992.
276. *Methyl Ethyl Ketone Peroxide in Dimethyl Phthalate (45:55)*, National Toxicology Program, Toxicity Report Series Number 18, NIH Publication 93-3341, United States Department of Health and Human Services, Washington D.C., Feb. 1993.
277. P. G. Watanabe, D. G. Pegg, J. D. Burek, H. O. Yakel, and L. W. Rampy, *A 90-Day Repeated Inhalation Toxicity Study of Cumene Hydroperoxide in Rats*, Toxicology Research Laboratory, EPA Document No. 868600016, Fiche No. OTS0510168, Dow Chemical USA, Midland, Mich., 1979.
278. P. Watts, *Food Chem. Toxicol.* **23**, 957 (1985).
279. J. E. Guillet and M. F. Meyer, *Ind. Eng. Chem. Prod. Res. Dev.* **1**, 226 (1962).
280. E. S. Shanley and J. R. Perrin, *Jet Propulsion*, 382 (1958).
281. *Safety of Organic Peroxides*, Technical Bulletin, Akzo Chemicals B.V., Devener, the Netherlands, 1990.
282. *Suggested Relative Hazard Classification of Organic Peroxides*, Technical Publication, Organic Peroxide Producers Safety Division. The Society of the Plastics Industry, Inc., New York, 1992.
283. ASTM Standard D 3278-89, "Flash Point of Liquids by Setaflash Closed-Cup Apparatus," ASTM, Philadelphia, Pa., 1989.
284. U. D. Wagle, G. W. Houston, A. J. Anderson, and O. L. Mageli, *Technical Data—Organic Peroxide Decomposition Characteristics*, Lucidol Division, Pennwalt Corp., Buffalo, N.Y., 1977.
285. J. J. de Groot, Th. M. Groothuizen, and J. Verhoeff, *IE&EC Proc. Des. Dev.* **20**, 131 (1981).
286. Th. M. Groothuizen, J. W. Hartgerink, and H. J. Pasman, in C. H. Buschmann, ed., *Loss Prevention and Safety Promotion in the Process Industries*, Elsevier Publishing Co., Amsterdam, the Netherlands, 1974, pp. 239-251.
287. J. L. C. van Geel, *Ind. Chem. Eng.* **58**, 24 (1966); J. J. de Groot and F. H. van der Elst, *Inst. Chem. Eng. Symp. Ser. No. 68*, 1 (1981).
288. *1972 Jahresberichte*, Bundesanstalt für Materialprüfung (BAM), Berlin, BRD.
289. H. G. Fisher and D. D. Goetz, *Determination of Self-Accelerating Decomposition Temperatures (SADTS) Using the Accelerating Rate Calorimeter (ARC)*, AIChE Meeting, Nov. 11-16, Union Carbide Chemicals and Plastics Co., Inc., S. Charleston, W. Va., 1990; M. W. Whitmore and J. K. Wilberforce, *J. Loss Prev. Process Ind.* **6**, 95 (1993); H. G. Fisher and D. D. Goetz, *J. Loss Prev. Process Ind.* **6**, 183 (1993).
290. *Recommendations on the Transport of Dangerous Goods, Tests and Criteria*, 2nd ed., printed on behalf of the United Nations by LABELMASTER, Div. of American Labelmark Co., Part III, *Tests and Criteria for the Classification of Organic Peroxides*, Chicago, Ill., 1990, pp. 211-315.
291. C. H. Bamford, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 13, Wiley-Interscience, New York, 1988, pp. 708-867.
292. *Ketone Peroxides*, Product Bulletin, Organic Peroxide Division, Elf Atochem North America, Inc., Philadelphia, Pa., Nov. 1991.
293. *Peroxyketals*, Product Bulletin, Organic Peroxides, Elf Atochem North America, Inc., Philadelphia, Pa., Apr. 1993.
294. *Diacyl Peroxides*, Product Bulletin, Elf Atochem North America, Inc., Philadelphia, Pa., Sept. 1992.
295. *Peroxyesters*, Product Bulletin, Elf Atochem North America, Inc., Philadelphia, Pa., May 1993.
296. *Tertiary Alkyl Hydroperoxides*, Product Bulletin, Atochem North America, Inc., Buffalo, N.Y., Nov. 1991.
297. *Initiators Product Bulletins*, AKZO Chemicals Inc., Dobbs Ferry, N.Y., 1993.
298. *Organic Peroxides Product Guide and other product bulletins*, Argus Division, Witco Corp., Marshall, Tex., 1992.
299. *Product Bulletins*, Laporte Organics-Aztec Peroxides-Interox-Peroxid-Chemie GmbH, Houston, Tex., 1993.
300. *Product Bulletins*, Reichhold Chemicals, Inc., Research Triangle Park, N.C., 1989.
301. *Product Bulletins*, Norac Co., Inc., Azusa, Calif., 1994.
302. *Product Bulletins*, Hercules Inc., Wilmington, Del., 1991.
303. *General Catalog—Peroxides & Specialty Chemicals*, Elf Atochem North America, Inc., Philadelphia, Pa., Oct. 1992.
304. **U.S. Pat. 4,092,470** (May 30, 1978), H. H. J. Oosterwijk and R. Torenbeek (to Akzona Inc.).
305. **U.S. Pat. 4,950,422** (Aug. 21, 1990), R. Torenbeek and W. F. Verhelst (to Akzona Inc.); **U.S. Pat. 5,260,390** (Nov. 9, 1993), R. Torenbeek and W. F. Verhelst (to Akzo NV).
306. **U.S. Pat. 3,058,926** (Oct. 16, 1962), J. Eichorn (to the Dow Chemical Company); **U.S. Pat. 3,058,929** (Oct. 16, 1962), J. W. Vanderhoff and A. K. Jahn (to the Dow Chemical Company).
307. Cr. I. Simionescu and co-workers, *Prog. Polym. Sci.* **12**, 1 (1986).
308. U.S. Pats. 3,706,818 (Dec. 19, 1972) and 3,839,390 (Oct. 1, 1974), O. L. Mageli and A. J. D'Angelo (to Pennwalt Corp.); **U.S. Pat. 4,079,074** (Mar. 14, 1978), J. Sanchez (to Pennwalt Corp.).
309. **U.S. Pats. 4,042,773** (Aug. 16, 1977) and 4,045,427 (Aug. 30, 1977), C. S. Sheppard and R. E. MacLeay (to Pennwalt Corp.).
310. **U.S. Pat. 4,822,883** (Apr. 18, 1989), T. N. Myers (to Pennwalt Corp.).
311. **U.S. Pat. 5,051,531** (Sept. 24, 1991), J. Sanchez (to Atochem North America, Inc.).
312. **U.S. Pat. 5,292,839** (Mar. 8, 1994), L. L. T. Vertommen, J. Meijer, and B. J. Maillard (to Akzo NV).
313. **U.S. Pat. 5,037,892** (Aug. 6, 1991), A. H. Hogt, J. Meijer, P. Hope, and J. Jelenic (to Akzo NV).
314. **U.S. Pat. 5,157,087** (Oct. 20, 1992), A. H. Hogt, J. Meijer, and J. P. J. Verdaan (to Akzo NV).

315. U.S. Pat. 5,202,386 (Apr. 13, 1993), A. H. Hogt, J. Meijer, P. Hope, and J. Jelenic (to Akzo NV).

Jose Sanchez
Terry N. Myers
Elf Atochem North America, Inc.

PERSULFATES.

See PEROXIDES AND PEROXIDE COMPOUNDS, INORGANIC

PERVAPORATION.

See HOLLOW FIBER MEMBRANES.

PESTICIDES

Developments in pesticide chemistry have been reviewed by scientists from around the world in proceedings of week-long congresses sponsored by the International Union of Pure and Applied Chemistry (IUPAC). The 6th International Congress on Pesticide Chemistry was held in 1986 in Ottawa, Canada (1), the 7th was in 1990 in Hamburg, Germany (2), and the 8th in 1994 was in Washington, D.C. (3). Additional information on pesticides can be found in publications based on symposia sponsored by the Agrochemicals Division of the American Chemical Society (4–24).

History

Devastation caused by pests has troubled both ancient and modern humans often changing the course of history. The bubonic plague in Europe and the great potato famine in Ireland were both caused by pests. In 1884, grasshoppers caused such a food shortage in the midwestern United States that a national disaster was declared.

Early attempts to control fungus on foliage relied on dusting with sulfur (qv), and Paris Green (cupric acetoarsenite) was applied on a large scale to kill the destructive Colorado potato beetle. Another highly toxic mixture used as a crop protectant was London Purple containing arsenic trioxide, aniline, lime, and ferrous oxide. By the early 1900s, chemicals used to control insect infestations included Bordeaux mixture (lime and copper sulfate) for downy mildew on grapes, nicotine sulfate for sucking insects, and lead arsenate and calcium arsenate for chewing insects. Increasing use of such toxic chemicals led to the passage of the Federal Insecticide Act in 1910, the first of many rules and regulations governing the sale and use of these pesticides, which were designated economic poisons at that time (see FUNGICIDES, AGRICULTURAL; INSECT CONTROL TECHNOLOGY).

Plants can also be pests that need to be controlled, particularly noxious weeds infesting food crops. Prior to 1900, inorganic compounds such as sulfuric acid, copper nitrate, sodium nitrate, ammonium sulfate, and potassium salts were used to selectively control mustards and other broadleaved weeds in cereal grains. By the early 1900s, Kainite and calcium cyanamid were also used in monocotyledenous crops, as well as iron sulfate, copper sulfate, and sodium arsenate. From 1915 to 1925, acid arsenical sprays, carbon bisulfate, sodium chlorate, and others were introduced for weed control use. Total or nonselective herbicides kill all vegetation, whereas selective compounds control weeds without adversely affecting the growth of the crop (see HERBICIDES).

When in World War II American farmers were called on to feed half the world population, research efforts led to development of more effective compounds. The first organochlorine insecticide, dichlorodiphenyltrichloroethane [50-29-3] (DDT) was soon followed by mixed isomers of benzene hexachloride (BHC), principally gamma-hexachlorocyclohexane [319-86-8] (lindane). Other insecticides included organophosphorus derivatives such as parathion [56-38-2] (O,O-diethyl O-4-nitrophenyl phosphorothioate) and malathion [121-75-5] (diethyl (dimethoxythiophosphorylthio)succinate), and carbamates such as carbaryl [63-25-2] (1-naphthyl methylcarbamate). Also recognized at this time were the plant growth regulating properties of compounds such as 2,4-D [94-75-7] (2,4-dichlorophenoxy acetic acid) and MCPA [94-74-6] (4-chloro-2-methylphenoxy acetic acid) (see GROWTH REGULATORS, PLANTS).

In early 1995, the U.S. Environmental Protection Agency (EPA) summarized its role in the regulation of pesticides:

All pesticide products created for use by homeowners and farmers in the U.S. must be registered by EPA. This process includes extensive testing to determine the toxicity of the product and its potential for threatening the health of people, wildlife, and the environment. Laws and regulations apply to all pesticides, including disinfectants, fungicides, insecticides, and weed-killers. When a pesticide is registered by EPA, the manufacturer is required to label it with specific instructions as to use, disposal, and special precautions. The label requires agricultural employers to provide their employees with many safety protections. If later scientific developments indicate unsuspected dangers, the registration can be suspended, canceled, or amended. EPA is expediting re-examination of the hundreds of pesticides registered during the Agency's early years using sound scientific standards. The Agency sets specific limits on pesticide residues in food, the limits depending on toxicity, and the quantity of those residues. Once EPA establishes the levels of pesticide that may remain on food, the Food and Drug Administration (FDA) and the Department of Agriculture (USDA) monitor the levels.

The original Insecticide Act of 1910 was replaced in 1947 by the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), which also regulated agents used to control fungi and animal pests such as rats and mice (see FUNGICIDES, AGRICULTURAL). In 1959, an amendment added herbicides (qv), nematocides, plant growth regulators (see GROWTH REGULATORS, PLANTS), defoliants, and desiccants. Another amendment in 1962 declared certain forms of plant and animal life, and viruses, to be pests under certain conditions. Principal revisions to FIFRA were enacted in 1972 and in 1988. Types of pesticides regulated under pesticide laws include the following (see also SOIL CHEMISTRY OF PESTICIDES).

acaricides	algicides	animal dips
antimicrobials	attractants	avicides
bactericides	baits	biological agents
defoliants	desiccants	disinfectants
fumigants	fungicides	herbicides

insect growth regulators	insecticides	ixodicides
larvicides	miticides	molluscicides
nematicides	pheromones	piscicides
plant growth regulators	repellents	rodenticides
safeners	seed protectants	soil sterilants
synergists	wood preservatives	wound protectants

The 1994 Pesticide Manual (25) has 725 entries for chemicals in use as pesticides, plus 560 additional chemicals believed to be no longer manufactured or marketed for crop protection use. The entries include many different classes of chemicals used as the active ingredients (AIs) in at least 35,000 formulated pesticide products worldwide. Many of these chemicals have more than one type of action from among those listed as regulated under pesticide laws. Each pesticide product used in the United States must be evaluated and registered by the EPA. Labeling must be approved for each use and each product must be registered in each state where it is sold and used. Pesticides are used for many diverse purposes, including the following.

- Crop Protection:* insects and weeds in fields, gardens, greenhouses, nurseries
- Disinfection:* bacteria in homes, hospitals, on medical and dental equipment
- Domestic Animals:* pests on dairy and beef cattle, sheep, goats, hogs, horses, poultry, and pets such as dogs, cats, and birds
- Forest Restoration:* conifer release; selective weed and brush control
- Fuel Preservation:* diesel oil, gasoline, heating oil, jet engine fuel
- Fumigation:* stored fruit and grain; quarantine of export and import crops
- Indoor Pests:* cockroaches, fleas, flies, lice, carpet beetles, clothes moths, silverfish, centipedes, millipedes, termites; mice, rats; mildew
- Outdoor Pests:* biting flies, fire ants, hornets, mosquitoes, ticks, wasps; mice, moles; snails, slugs; mildews, molds, mosses
- Pests in Aquatic Sites:* weeds clogging navigable streams, infesting recreational areas; predator eels and fish; zebra mussels
- Post-Harvest Treatment:* fresh produce during transportation, distribution, storage
- Seed Treatment:* prevent spoilage, premature germination, sprout growth
- Transport Equipment:* insects and rodents carried by trucks, trains, ships, aircraft
- Tree Preservation:* gypsy moths, caterpillars, borers, leaf miners, mites; wounds from pruning or other damage to bark
- Turf Protection:* insects and weeds in lawns, parks, golf courses; moles
- Vector Control and Plagues:* rodents, mosquitoes, tsetse flies, grasshoppers, locusts
- Vegetation Control:* rights-of-way along fence rows, roads, highways, railroads, utility lines, pipelines
- Water Purification:* reservoirs, swimming pools, cooling towers
- Wood Preservation:* construction lumber, fence posts, railroad ties, utility poles

Production

According to the EPA, more than 4.5 × 10⁵ t of pesticide AIs and related products were used in the United States in 1988, consisting mainly of herbicides (310,000 t), insecticides (120,000 t), and fungicides (59,000 t). Five years later (1993), the total volume used in the United States was unchanged, although the amounts of herbicides and insecticides had declined somewhat (26). This decrease was partly the result of discontinuation of many older products that had been used at high rates (on the basis of AI, ~50 kg ha (<50 lb acre, and development of several new classes of chemicals effective at very low rates (in AI, g acre or g ha). In addition, EPA approved a number of biological agents to replace some of the more toxic chemicals. Such biological agents, often called natural insecticides, are derived from living organisms (9,10,27). For example the bacterium *Bacillus thurengiensis* (B.t) produces a toxin that injures the gut of insect larvae. Different strains of B.t. can be selected to control various insects in home gardens, agriculture, and forestry. Many urban areas rely on spraying with B.t. *israelensis* to control mosquitoes.

Table 1 compares pesticide markets in the United States and worldwide. In 1993, the total volume used in the United States was 24° of the total active ingredients used worldwide, whereas the total spent in the United States was 34° of that expended worldwide. Although the United States used proportionately less insecticides (15° of total worldwide), insecticide expenditures were 32° of the total worldwide primarily because of the higher insecticide costs. These costs are estimated to average more than twice as high in the United States as worldwide, possibly because many of the older less expensive insecticides are no longer permitted for use in the United States. On the other hand, the average cost of fungicides was lower in the United States than worldwide. A primary factor in raising prices in the United States has been the need for manufacturers to generate extensive additional data to support reregistration of all active ingredients (AIs) contained in products registered in the United States prior to November 1984 (28). These new data must meet the standards for registration of new pesticides.

Almost 80° of pesticide sales are in developed countries (29). In addition to the 24° sold in the United States in 1993, 28° of the total was sold in Western Europe, 25° in the Far East, 8° in Eastern Europe and the former USSR, 7° in Latin America, and 7° in the rest of the world. Brazil is the world's fourth largest user of pesticides after the United States, Japan, and France. Overall pesticide usage in agriculture, in terms of amounts applied per acre or hectare, has been much greater in Japan, Europe, Taiwan, and Australia than in the rest of the world. China, however, is also a significant user and usage in Vietnam and Indonesia is expanding rapidly (30). In 1988 the intensity of pesticide use was high in Europe, ranging from 18.5 kg ha in the Netherlands, 5.8 in the United Kingdom, 4.5 in France, 4.2 in West Germany, 2.6 in Denmark, and 1.5 in Sweden, compared to 1.8 kg ha in the United States and 0.9 in Canada (29). Some governments have introduced legislation to reduce pesticide use by 50° or more by the year 2000, primarily because of concern about potential contamination of drinking water (31).

In developing countries, 90° of pesticide use is on agricultural crops, particularly on cotton (qv) and rice (see WHEAT AND OTHER CEREAL GRAINS). The remaining 10° is used in vector control of human diseases (see ANTIPARASITICS; ANTIVIRAL AGENTS) (30). About 1.9 × 10⁹ people are at risk from malaria in 102 countries, particularly in Africa, southeast Asia, and the western Pacific. DDT was formerly used for indoor residual application on a wide scale, but resistance to this and other organochlorine insecticides has been reported in 51 species of mosquitoes. The replacement organophosphate and carbamate insecticides are generally not as effective as DDT and are considerably more expensive. River blindness (onchocerciasis) is caused by a nematode transmitted by blackflies that can be controlled by applying larvicides to rivers in affected zones. However, larvicides may cause adverse effects in fish and aquatic invertebrates. Schistosomiasis is caused by a trematode transmitted by aquatic snails that can be controlled by molluscicides, but these agents may also present hazards to aquatic organisms. Trypanosomiasis is caused by a protozoan transmitted by the tsetse fly, controlled in the past by ground spraying the bush habitats of the flies with residual insecticides such as DDT and dieldrin, but these agents have been shown to cause adverse effects in wildlife (30).

Table 1. Estimates of U.S. and World Use of Conventional Pesticides, 1993^a

Parameter	Herbicides	Insecticides	Fungicides	Other	Total
quantity, t (°)					
U.S.	280,000 (57)	112,000 (23)	59,000 (12)	38,000 (8)	490,000 (100)
world	960,000 (47)	739,000 (36)	243,000 (12)	105,000 (5)	2,000,000 (100)
quantity, U.S. world, °	29	15	24	36	24
total cost, U.S.\$ × 10 ⁶ (°)					
U.S.	4,756 (56)	2,550 (30)	584 (7)	594 (7)	8,484 (100)

world	11,700 (46)	7,900 (31)	4,130 (16)	1,550 (6)	25,280 (100)
cost, U.S. world, %	41	32	14	38	34
av cost, U.S.\$ kg AI (%)					
U.S.	16.87 (98)	22.70 (131)	9.79 (57)	15.73 (91)	17.23 (100)
world	12.21 (99)	10.69 (86)	16.96 (137)	14.83 (120)	12.36 (100)
av cost, U.S. world, %	138	212	58	106	140

^a Adapted from Ref. 26.

Only very large companies can afford the high costs of developing pesticides, estimated to be more than U.S.\$70 × 10⁶ for a successful new active ingredient (29,32). This cost includes research on the ~20,000 chemicals investigated for each one that is eventually registered and achieves a return on investment in manufacturing facilities and sales promotion. The market is a competitive one (33). Similarly, only large companies are able to support the research needed to generate the data required for reregistration of existing pesticide products in the United States (28).

By 1989 about 75% of the world pesticide market was dominated by 10 firms (34). Extensive consolidation of the agrochemical industry came about between 1986 and 1993, during which time at least 40 acquisitions or mergers to form joint ventures occurred (35). Of the top 20 companies in pesticide sales worldwide in 1993 (Table 2), nine had headquarters in Europe, six in the United States, and the remaining five were based in Japan (36). Most of these companies also have global interests, having sales offices in many developed and developing countries. In 1994, the agrochemical portions of Hoechst and Schering (including NorAm) became AgrEvo, joined later by Roussel Uclaf moving this new company up to second place. Similarly, American Cyanamid acquired Shell and moved up to eighth place before being acquired by American Home Products in December 1994.

Table 2. Ranking of Principal Agrochemical Companies by Sales in 1993^{a,b}

Rank	Company	U.S.\$ × 10 ⁶	Rank	Company	U.S.\$ × 10 ⁶
1.	Ciba-Geigy	2788	11.	Sandoz	890
2.	Du Pont	2014	12.	Schering	703
3.	Monsanto	1936	13.	Shell	650
4.	Zeneca	1916	14.	FMC	475
5.	Bayer	1791	15.	Kumiai	443
6.	Rhône-Poulenc	1745	16.	Ishihara	425
7.	DowElanco	1606	17.	Sumitomo	423
8.	Hoechst	1355	18.	Sankyo	416
9.	BASF	1119	19.	Rohm & Haas	409
10.	American Cyanamid	1106	20.	Nihon Nohyaku	407

^a Ref. 36.

^b Total sales by the top 20 agrochemical companies was U.S.\$22.6 × 10⁹ in 1993.

Technology

Pesticide use has undergone many changes since toxic products containing arsenic and mercury were dusted indiscriminately in attempts to control pests infesting crops or homes. Advances in pesticide technology have occurred in discovery, production, formulation, and application. The relatively few pesticides available to previous generations of farmers have been replaced by hundreds of highly active agrochemicals targeted to control specific pests in specific sites under specific conditions, such as selective inhibition of certain enzymes in plants and insects.

Discovery. The traditional approach to new pesticide discovery was to make intuitive changes in the substituents on a promising primary chemical structure. Initially, materials from any source were subjected to screening for biological activity as insecticides, herbicides, or fungicides. Experience showed that, on average, only one in about 20,000 chemicals tested achieved commercialization (33). The challenge to speed up discovery of novel pest control agents effective at low application rates, highly selective in biologic action, and having demonstrably low impacts on the environment, is being met. Optimization in selection of discovery pathways has been enhanced through use of tools such as quantitative structure activity relationship (QSAR) and computer-aided molecular design (CAMD) (see MOLECULAR MODELING (SUPPLEMENT)). Application of this technology has been reviewed by several experts (37–41).

Manufacturing. Early pesticides (Table 3) were simple compounds, often easy to make. Some of these have a high mammalian toxicity and present unacceptable hazards to farmers and other agricultural workers. In contrast, the manufacture of new chemical classes of pesticides having complex structures generally requires multistep synthesis processes, any of which can lead to side products or impurities (4–7,42). Also, concern about the fate of bioactive chemicals introduced into the environment has led to strict regulations on release of vapors into air or of manufacturing waste into effluent water, as well as for proper disposal of containers and wastes from pesticide use (8).

Table 3. Examples of Simple Early Pesticides

Compound	CAS Registry Number	Molecular formula	Pesticide class
formaldehyde	[50-00-0]	HCHO	bactericide
bromomethane	[74-83-9]	CH ₃ Br	fumigant
dimethylarsinic acid	[75-60-5]	(CH ₃) ₂ AsO ₂ H	herbicide
dalapon	[75-99-0]	CH ₃ Cl ₂ CO ₂ H	herbicide
trichloroacetic acid	[76-03-9]	Cl ₃ CCO ₂ H	herbicide
acrolein	[107-02-8]	CH ₂ =CHCHO	herbicide
2-phenylphenol	[90-43-7]	C ₁₂ H ₁₀ O	fungicide
biphenyl	[92-52-4]	C ₁₂ H ₁₀	fungicide
diphenylamine	[122-39-4]	C ₁₂ H ₁₁ N	fungicide
mercuric oxide	[21908-53-2]	HgO	fungicide
mercurous chloride	[7546-30-7]	Hg ₂ Cl ₂	fungicide
sodium fluoride	[768-49-4]	NaF	insect bait

Some of the newer compounds may contain both saturated and unsaturated rings, heteroatoms such as oxygen, nitrogen, or sulfur, and halogen substituents. Others, such as synthetic pyrethroids, may have one or more chiral centers, often needing stereospecific methods of synthesis or resolution of isomers (42). Table 4 lists examples of some more complex compounds. Structures are shown in Figure 1 (25).

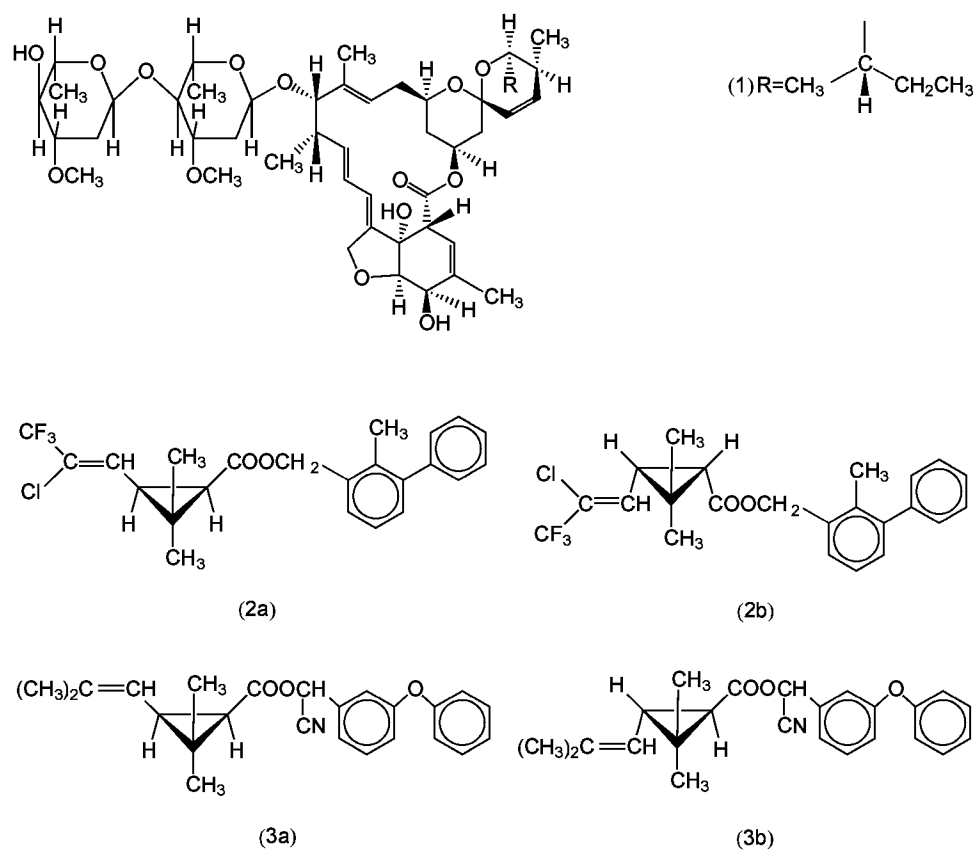


Fig. 1. Structures of more complex pesticides. See Table 4. Structures (2a) and (2b) are the (Z)-(1R)-cis and (Z)-(1S)-cis isomers, respectively; (3a) and (3b) are the (1R)-cis and (1R)-trans isomers, respectively.

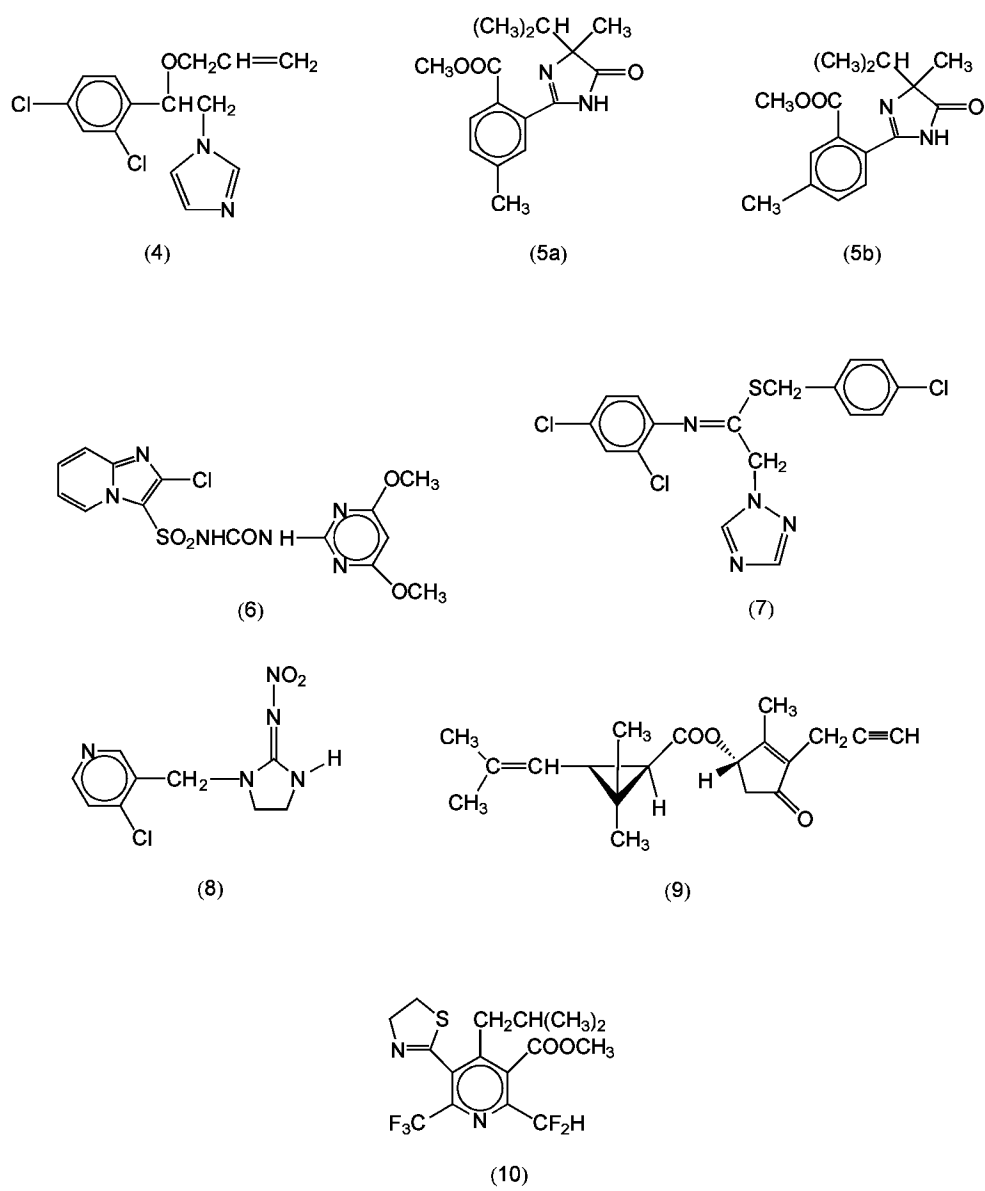


Table 4. Examples of Complex Pesticides^a

Compound	CAS Registry Number	Molecular formula	Structure number ^b	Pesticide class
abamectin ^c	[71751-41-2]	C ₄₈ H ₇₂ O ₁₄	(1)	acaricide
bifenthrin	[82657-04-3]	C ₂₃ H ₂₂ ClF ₃ O ₂	(2a) plus (2b)	acaricide
cyphenothrin	[39515-40-7]	C ₂₄ H ₂₅ NO ₃	(3a) plus (3b)	insecticide
imazalil	[73790-28-0]	C ₁₄ H ₁₄ Cl ₂ N ₂ O	(4)	fungicide
imazamethabenz	[81405-85-8]	C ₁₅ H ₁₈ N ₂ O ₃	(5a) plus (5b)	herbicide
imazosulfuron	[122548-33-8]	C ₁₄ H ₁₃ ClN ₄ O ₅ S	(6)	herbicide
imibenconazole	[86598-92-7]	C ₁₇ H ₁₃ Cl ₃ N ₄ S	(7)	fungicide
imidacloprid	[105827-78-9]	C ₉ H ₁₀ ClN ₅ O ₂	(8)	insecticide
prallethrin	[23031-36-9]	C ₁₉ H ₂₄ O ₃	(9)	insecticide
thiazopyr	[117718-60-2]	C ₁₁ H ₁₇ F ₅ N ₂ O ₂ S	(10)	herbicide

^a Ref. 25.
^b See Fig. 1.
^c Abamectin was first isolated from seeds of the neem tree as a mixture of (1), as shown in Fig. 1, plus (1, R = CH(CH₃)₂).

Formulation, Packaging, and Distribution. Advanced technology is needed to formulate pesticides to meet the needs of farmers and commercial applicators who must operate under regulatory constraints for protection of the environment. Basic producers of pesticide active ingredients (AIs) can formulate their own products or have the products formulated by secondary companies. Some formulators also manufacture pesticides that are no longer covered by patents, or import AIs from foreign producers. However, each AI and pesticide product formulated in the United States must be registered by the EPA before it can be distributed and sold. It must have approved labeling for each recommended use, and must also be registered by each state where it is sold and used.

Much progress has been made. Historically compounds containing arsenic and mercury were applied as dusts in orchards and cotton fields (21). Wettable powders and emulsifiable concentrates were developed that could be applied by spraying, as were granular products suitable for application with equipment used for solid fertilizers. More recently, dry flowable formulations have been developed. These are easier to use, safer for mixer loaders to handle, and generally perform better than wettable powders. Water dispersible granules are also easier to measure than powders and can be applied using conventional spraying equipment (43–46). Newer innovations include premeasured amounts of concentrates in water-soluble pouches packaged in recyclable paper that applicators can handle without contacting the product (44). Some companies have also developed proprietary encapsulated formulations and slow-release products that allow pesticides to become activated in the soil by trigger mechanisms such as moisture or temperature (see CONTROLLED RELEASE TECHNOLOGY, AGRICULTURAL).

Inert ingredients include solvents, emulsifiers, surfactants (qv), dispersants, stabilizers, preservatives, sequestrants, and other substances. These are not AIs, but aid in ensuring consistent action of one or more AIs in a formulated product. The products must remain stable during storage and distribution under a variety of environmental conditions, ranging from extreme heat in southern regions to subzero temperatures in northern areas. Containers must not rust or leak, must withstand rough handling, and must not collapse when stacked in warehouses. The openings must be childproof and liquid contents should not gurgle or splash when being poured. Totally enclosed systems were introduced by some companies for highly toxic products so the contents could be transferred directly to the mixing tanks, but the special equipment required is often not available at the general user level (44). Since 1990, closed containers have been widely used for bulk packaging units for new potent herbicides.

In 1987, EPA issued a policy statement regarding inert ingredients in formulated pesticide products, and in 1992, published a list of inerts in four categories according to the degree of toxicological concern. Labeling on any product containing a List 1 inert had to be revised to show the statement: "This product contains the toxic inert (name of inert)." Also, the EPA strongly encouraged registrants to substitute or remove from their products any List 1 or List 2 inert ingredients and to submit revised confidential statements of formula to amend these registrations. Updated Lists 1 and 2 are given in annual editions of the *Farm Chemicals Handbook* (see *General References*).

List 2 inerts are considered to be potentially toxic and at high priority for testing. List 3 includes about 800 inert ingredients that have no basis for being on Lists 1, 2, or 4, and List 4 contains inerts which are generally regarded as safe (GRAS). This latter group includes about 300 inert ingredients, such as clays (qv), cookie crumbs, corn cobs, etc.

As more effective pesticides were discovered, application rates dropped from kilograms or liters of formulated products per acre to grams of active ingredient per hectare (g/ha of AI). Highly active, low volume products are easier to ship to dealers, take up less storage space until sold to customers, and are easy to deliver to farmers. The smaller quantities needed can be supplied in 0.5–10 L (pint to 2.5 gal) plastic jugs that can be triple rinsed easily as required before proper disposal, or in some cases can be returned to the manufacturer for reuse (8). On the other hand, the large quantities needed for older, less active products were generally distributed in bulky metal drums, such as those formerly disposed of improperly in ditches, along streams, or in domestic landfills. Bulk shipments of tank-car lots can be made to large-scale commercial applicators having approved diked platforms for mixing and loading, such as for herbicide products to be applied by aircraft.

Application of Pesticides. Older, less active pesticides were often applied using backpack tanks and hand-held wands to direct the spray onto target weeds and brush. For somewhat larger projects, the sprays were applied from tractor-drawn rigs with booms extending over several rows of field crops or onto adjacent rights-of-way. Tractor-drawn mistblowers dispense insecticides into the foliage canopy of trees in fruit and nut orchards (see NUTS), but this procedure is hazardous to the driver unless the tractor cab is completely enclosed, air-conditioned, and equipped with charcoal filters on air intake vents. Formulations that provide large droplets are necessary for herbicides applied by aircraft, including helicopters, under low wind conditions to avoid drift from the target area. Fogging sprays can be used to control insect pests such as mosquitoes in urban and recreational areas, and destructive insects like gypsy moths in forests. Very dilute ready-to-use products are packaged in aerosol spray cans and are registered for use around homes and institutions.

More efficient application systems have been developed, including longer spray booms to cover a greater area per pass and installation of movable or permanent tanks to dispense the prepared sprays. More sophisticated equipment is essential for accurate and even application of highly effective pesticides, particularly for products formulated for low volume or ultralow volume (ULV) spraying. All equipment must be carefully calibrated to ensure that the rates of application are the same from nozzles located at the end of a boom as from those near the spray rig. In some cases, prescription amounts can be applied according to need, based on computer calculations using satellite systems to outline contours and provide information on moisture and temperature conditions in different areas of large fields (44). Some products have also been approved by the EPA for drip application in irrigation water, or from central pivot sprayers above artesian wells on the circular fields of the arid Great Plains area in the United States.

Regulation

Pesticides are more closely regulated than other chemicals because pesticides are intentionally applied in the environment, often repeatedly at relatively high rates. In the United States, pesticides are regulated under the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), and residues from uses of pesticides in food or feed crops are regulated under Sections 408 and 409 of the Federal Food, Drug and Cosmetics Act (FFDCA).

FIFRA has been amended a number of times since it was first enacted in 1947. It was originally administered by USDA. Largely rewritten in 1972, FIFRA was transformed from a limited consumer protection act into a full environmental law. The EPA, created in 1970, was given the task of

implementing this law, including registration of all new pesticides and reregistration of all old pesticides. The EPA also took over responsibility from the FDA for setting maximum residue levels (tolerances) for pesticide residues in foods and feeds. Section 408 of FFDCA pertains to residues in raw agricultural commodities (RACs) and processed foods, except when the residue in the processed food exceeds the Section 408 tolerance for the parent RAC. Residues that concentrate in processed food are deemed unintentional food additives (qv), and are granted higher tolerances under provisions of Section 409 of FFDCA. Section 409 includes the outdated Delaney Clause passed by Congress in 1958, which provides that "no additive shall be deemed to be safe if it is found to induce cancer when ingested by man or animal, or if it is found, after tests which are appropriate for the evaluation of the safety of food additives, to induce cancer in man or animal" (47). Furthermore, under Section 409, no consideration can be given to the proven benefits from continued use of pesticides that cause tumors, ie, are oncogenic; regardless of the insignificance of the risk estimated by ultraconservative extrapolations from studies conducted in animals given maximum tolerated doses daily over the entire lifespans. The inconsistency between regulation of pesticide residues in raw foods and processed foods has been the subject of intense debate for many years (48) and is not resolved as of this writing (mid-1995).

Initially, the EPA's Office of Pesticide Programs (OPP) focused its efforts on canceling DDT and instituted a tedious four-step review process called Rebuttable Presumption Against Registration (RPAR) for in-depth evaluations of other persistent organochlorine insecticides, and of other pesticides deemed to present an unacceptable hazard to human health or the environment. RPAR has evolved into a formal special review process in which the EPA must consider a pesticide's risks and benefits, which allows for public input to the decision-making process, and which can result in a decision to cancel, restrict, or continue the pesticide uses in question. By mid-1994, the EPA had conducted special reviews on more than 100 pesticides or groups of closely related pesticides (49), had revoked tolerances for residues of 53 canceled or withdrawn pesticides, and had proposed revoking tolerances for some uses of 31 additional pesticides classified as probably or possibly oncogenic in animals (50).

Requirements for Pesticide Registration in the United States. Pesticide registration decisions are based primarily on EPA evaluations of test data provided by applicants to ensure that, when used according to label directions, the pesticide does not cause unreasonable adverse effects to human health or the environment. Testing is needed to show whether the pesticide has the potential to cause adverse effects to humans, wildlife, fish, and plants, including endangered species. Potential human risks, which are identified using laboratory tests in animals, include acute toxic reactions such as poisoning and skin and eye irritation, as well as possible long-term effects such as cancer, birth defects, and reproductive disorders. Data on the fate of pesticides in the environment are also required so that scientists in OPP can determine, among other things, whether a pesticide poses a threat to groundwater or surface water (lakes, rivers, and streams). Extensive analytical studies are also required to establish maximum residue levels anticipated from recommended uses in food or feed crops.

FIFRA Sections 3 and 4 pertain to registration and reregistration of pesticides, with clearly defined data requirements as outlined in Title 40 of the *U.S. Code of Federal Regulations* (51). About 120 different studies are listed, most of which are to be done on technical-grade active ingredients (TGAI)s. Some must also be done on formulated products containing inert ingredients. All studies must be done according to a published series of Pesticide Assessment Guidelines (PAGs) and must meet supplementary requirements delineated in various Data Reporting Guidelines (DRGs), Standard Evaluation Procedures (SEPs), and additional guidance documents available from the EPA (see *General References*). All studies must be conducted under conditions to meet Good Laboratory Practice (GLP) standards (40 CFR 160), regardless of whether carried out in-house by registrants or by outside testing facilities. The group of tests that must be performed for each pesticide depends on how that pesticide is to be used. For example, if a pesticide is not used on food or feed crops, extensive residue and metabolism tests in plants and domestic animals might not be required. Similarly, if a pesticide is not used in field crops nor on other extensive outdoor areas, all the environmental fate studies might not be required.

Product Chemistry. Data and information from product chemistry studies (40 CFR 158.150–190 and Subdivision D Guidelines) are used by the EPA primarily to establish the composition of each manufacturing and user product as commercially produced. Product composition is reported in the Confidential Statement of Formula (CSF) which identifies and gives the amounts of the AI(s), intentionally added inert ingredients, and in some cases the impurities contained in each product. This basic information is needed to assess toxicity to humans and hazards to the environment resulting from use of the product, and to assess the physical and chemical hazards such as corrosiveness, explosiveness, and flammability. Analysis of the technical-grade active ingredient (TGAI) must account for all components present at 0.1% or more, and should account for >98% of the product used for manufacturing the product. The composition of all commercial products must be within certified limits set at the time of registration.

Residue Chemistry. Residue chemistry studies must be carried out according to 40 CFR 158.240 Subdivision O Guidelines. The data are used by the EPA to estimate the exposure of the general population to pesticide residues in food, and for setting and enforcing tolerances for pesticide residues in raw agricultural commodities or processed fractions that can be used as food or feed. Uses that can result in residues in meat, milk, poultry, or eggs are also considered to be food uses. Samples to be analyzed are obtained from extensive field studies in which the product is applied at the maximum recommended rate and frequency, using the shortest interval between the last treatment and harvest for each crop on which that product is to be used. The analytical method(s) must account for the total residue, including significant metabolites, and must be sensitive to a quantitation limit in the low parts per billion (ppb) range for any item used as food for humans (18,52–56). The method must have been validated by an outside laboratory and by an EPA laboratory, and must be suitable for enforcement purposes when used in monitoring studies conducted by the FDA and USDA.

Section	Topic
171-2	Chemical identity
171-3	Directions for use
171-4(a),(b)	Nature of residue in plants, livestock
171-4(c),(d)	Residue analytical method (plants, animals)
171-4(e)	Storage stability
171-4(f)–(h)	Magnitude of the residue in potable water, fish, irrigated crops
171-4(i)	Magnitude of the residue in food handling establishments
171-4(j)	Magnitude of the residue in meat milk poultry eggs (feeding dermal treatment)
171-4(k)	Crop field trials (for each crop use, in each geographic location)
171-4(l)	Magnitude of the residue in processed food feed
171-5	Practical methods for reduction of residues (by washing, peeling, cooking, etc)
171-6	Proposed tolerance (in each crop and crop by-product)
171-7	Reasonable grounds in support of tolerance petition
171-13	Analytical reference standard(s)

Subdivision O guidelines for residue chemistry data were originally published by the EPA in 1982. These have been supplemented to improve the rate of acceptance by EPA reviewers of the many reports submitted by registrants in support of tolerances for pesticides in foods. The residue chemistry studies most frequently rejected include metabolism in plants, food processing (qv) studies, and studies on storage stability of residues in field samples (57). All tolerances (maximum residue levels) established under FIFRA are listed in 40 CFR under Sections 180 for individual pesticides in on raw agricultural commodities, 180 for exemptions from tolerances, 185 for processed foods, and 186 for animal feeds.

Environmental Chemistry. Requirements for data on pesticides in the environment include both laboratory and field studies. The purpose of these studies is to identify and assess the potential hazards associated with each use of a pesticide in the environment in which it is to be used (20). These studies, governed by 40 CFR 158.290 and Subdivision N Guidelines, are generally conducted by or on the behalf of the basic manufacturer, using technical-grade chemical (TGAI), typical product, or a radioactively labeled analytical-grade chemical for studies where a material balance is needed. Studies include the following

Section	Topic
---------	-------

160-5	Chemical identity
161-1	Hydrolysis studies
161-2,3,4	Photodegradation studies in water, soil, air
162-1	Aerobic soil metabolism studies
162-2	Anaerobic soil metabolism studies
162-3	Anaerobic aquatic metabolism studies
162-4	Aerobic aquatic metabolism studies
163-1	Leaching and adsorption desorption studies
163-2,3	Volatility studies (laboratory, field)
164-1	Terrestrial field dissipation studies
164-2	Aquatic field dissipation studies: soil and sediment
164-3	Forest field dissipation studies
164-4	Dissipation studies for combination products and tank mix uses
164-5	Long-term soil dissipation studies for products
165-1	Accumulation studies in confined rotational crops
165-2	Accumulation studies in field rotational crops
165-3	Accumulation studies in irrigated crops
165-4	Bioaccumulation studies in fish (laboratory)
165-5	Bioaccumulation studies in aquatic nontarget organisms (field)
166-1	Groundwater monitoring study (small-scale prospective)
166-2	Groundwater monitoring study (small-scale retrospective)
166-3	Groundwater monitoring study (large-scale retrospective)

Controlled laboratory studies are required to examine the persistence, mobility, and potential for accumulation of a pesticide and its primary degradates. Persistence studies examine the behavior of a pesticide as it interacts with water, soil, air, sunlight, and microorganisms. Mobility studies attempt to predict the potential of the pesticide to volatilize into the atmosphere, move into ground or surface waters, or bind to the soil. Accumulation studies examine the potential for a pesticide to accumulate in rotational crops and fish. These studies are designed to help identify which dissipation processes are likely to occur when the pesticide is released into the environment and to characterize the significant degradates likely to result from these processes. From the results of these studies, EPA scientists develop a preliminary, qualitative assessment of environmental fate and transport. The data are then used to design and or trigger appropriate field studies and to provide parameters needed in simulation modeling of the environmental fate of pesticides (58) (see SOIL CHEMISTRY OF PESTICIDES).

Field studies are required to provide a more realistic picture of the dissipation of the parent compound and those degradates determined to be significant. Under field conditions pesticides are exposed simultaneously to the individual dissipation processes that were examined separately in the laboratory studies. Thus, in field studies, some dissipation processes may be altered due to competition and interaction. Requirements for spray drift data were outlined in draft Subdivision R, but the EPA agreed that data generated on a generic basis by an industry consortium could represent the potential for drifting of individual pesticides.

Data from field and laboratory studies are then integrated to characterize the persistence and transport of the pesticide and its degradates in the environment, and to develop a quantitative environmental assessment (59). Environmental concentrations of the pesticide in different media under various pesticide application and site scenarios are also calculated using computer modeling. These estimates of exposure are used in conjunction with toxicity data to assess the risks to nontarget species associated with the use of the pesticide. Computed risks are used by the EPA to determine the required degree of regulatory action which can include label advisories, use restrictions, denial of registration for a new pesticide, or suspension or cancelation of a registered pesticide. If the data warrant, a pesticide can also be placed in the Special Review process (40 CFR 154) to undergo a more extensive examination of specific problems uncovered during reviews.

Many of these studies require tremendous expenditure of time and effort, and should not be initiated until after consultation with scientists in the EPAs Environmental Fate and Groundwater Branch (20). As for residue chemistry data requirements, the environmental fate guidelines have been supplemented by DRGs and SEPs (58). Although the rejection rate for all environmental fate studies dropped from 54% for pre-1986 studies to 28% for studies submitted after 1988 (49), many problems still remain (60). Critical evaluations of the environmental fate guidelines were conducted by task forces consisting of scientists from the EPA, USDA, industry, and academia, who recommended how laboratory and field studies should be conducted to provide a more scientific basis for environmental risk assessment (20). Available data for many pesticides can be combined in computer modeling systems to predict the environmental behavior of individual pesticides without having to conduct costly and time-consuming field studies that provide only limited additional information for those pesticides (59). The EPA recognizes the need for revising these guidelines (61), and for harmonizing U.S. requirements with those in the European Union (62) and those developed by the international Organization for Economic Cooperation and Development (OECD).

Hazard Evaluation.

Humans and Domestic Animals. Data from toxicology studies are used to evaluate hazards to humans from the use of pesticides (40 CFR 158.340 and Subdivision F Guidelines).

Section	Topic
81-1	Acute oral toxicity (rat)
81-2	Acute dermal toxicity (rabbit)
81-3	Acute inhalation toxicity (rat)
81-4	Primary eye irritation (rabbit)
81-5	Primary dermal irritation (rabbit)
81-6	Dermal sensitization (guinea pig)
81-7	Acute delayed neurotoxicity (hen)
82-1(a),(b)	Subchronic 90-d feeding (rodent, nonrodent usually dog)
82-2	Subchronic 21-d dermal toxicity (albino rabbit)
82-3	Subchronic 90-d dermal toxicity (rat)
82-4	90-d Inhalation (rat)
82-5(a)	28-d Delayed neurotoxicity (adult hen)
82-5(b)	90-d Neurotoxicity in the mammal (rat preferred)
83-1(a),(b)	2-yr Chronic feeding study (rat; nonrodent, dog)
83-2(a)	2-yr Oncogenicity study (rat)
83-2(b)	Oncogenicity study (lifetime) (mouse)
83-3(a),(b)	Teratogenicity (development toxicity) (rat, rabbit)
83-4	Reproduction study (two generations) (rat)
83-5	Combined chronic feeding oncogenicity study (rat)
84-2(a)	Gene mutations
84-2(b)	Structural chromosome aberrations
84-2(c)	Other genotoxic effects

85-1	General metabolism (biotransformation)
85-2	Dermal penetration
86-1	Domestic animal safety

The acute studies provide information on health hazards likely to arise soon after, and as a result of a single exposure. These data are used to classify the pesticides as highly toxic (Category I), toxic (Category II), moderately toxic (Category III), or low in toxicity (Category IV), all of which require precautionary labeling using signal words such as Danger and Poison, Warning, or Caution (40 CFR 156.10). Acute data are used to set reentry intervals for farmworkers, to require protective clothing for applicators (40 CFR 170), and determine the need for child-resistant packaging (40 CFR 157). These data also provide information for establishing appropriate dose levels in the subchronic and other studies, and provide initial information on the mode of toxic action(s) of a substance.

Subchronic testing provides information on health hazards that might arise from repeated exposure to a chemical over a limited period of time. These studies can identify target organs and accumulation potential, and are also useful in selecting the maximum tolerated dose (MTD) levels for chronic studies and for establishing safety criteria for human exposure. The chronic toxicity studies are intended to determine the effects of a substance in mammalian species following prolonged and repeated exposure, and should detect effects which have a long latency period or are cumulative. The daily doses given to the animals should span a level that causes no observed effect (NOEL), the lowest effect level (LEL), and the maximum tolerated dose (MTD), just below the level that causes lethal effects. In oncogenicity studies in rats and mice, the test animals are observed for the development of neoplastic lesions (cancer) or benign tumors during or after feeding doses including the MTD over their normal lifespans (63,64).

The developmental testing sequence is designed to determine the potential of the test substance to induce structural and or other abnormalities in the fetus as a result of exposure to the mother during pregnancy. The two-generation reproduction test is designed to provide information concerning the general effects of a test substance on gonadal function, estrus cycles, mating behavior, conception, parturition, lactation, weaning, and the growth and development of offspring. The data generated from these studies are the NOEL, LEL, and the developmental toxicity potential, as well as the margins of safety for dietary and nondietary exposure.

The purpose of mutagenicity testing is to assess the potential for the pesticide to affect the qualitative or quantitative integrity of the mammalian cell's genetic components. The assays are selected to detect the capacity of a chemical to alter genetic material in cells, to determine the relevance of these mutagenic changes to mammals, and when mutagenic potential is detected, to incorporate these findings in the assessment of heritable effects, carcinogenicity, and possibly other endpoints. Other special tests are required as needed, such as delayed neurotoxicity for organophosphate and thiophosphate pesticides, and dermal penetration of substances for which the assumption of 100% absorption does not produce an adequate margin of safety for workers exposed to the substance.

Wildlife and Aquatic Organisms. Studies required to provide ecological effects data to determine the toxicological hazards of pesticides to various terrestrial and aquatic nontarget organisms are summarized (40 CFR 158.490 and Subdivision E Guidelines).

Section	Topic
71-1(a)	Acute avian oral toxicity (LD ₅₀) in bobwhite quail or mallard duck
71-1(b)	Acute avian oral toxicity (LD ₅₀) in bobwhite quail or mallard duck (using typical product)
71-2(a)	Acute avian dietary toxicity (LC ₅₀) in bobwhite quail
71-2(b)	Acute avian dietary toxicity (LC ₅₀) in mallard duck
71-3	Wild mammal toxicity test
71-4(a)	Avian reproductive toxicity in bobwhite quail
71-4(b)	Avian reproductive toxicity in mallard duck
71-5(a)	Simulated terrestrial field study
71-5(b)	Actual terrestrial field study
72-1(a),(b)	Fish toxicity in bluegill sunfish
72-1(c),(d)	Fish toxicity in rainbow trout
72-2(a),(b)	Invertebrate toxicity freshwater LC ₅₀ (daphnia preferred)
72-3	Toxicity to estuarine and marine organisms (six tests)
72-4(a)	Early life stage in fish
72-4(b)	Life cycle in aquatic invertebrates (daphnia mycid)
72-5	Fish life cycle study
72-6	Aquatic organism accumulation study
72-7(a),(b)	Field tests for aquatic organisms, simulated, actual

A risk characterization for a pesticide use is performed by comparing these effects data with data on environmental fate and exposure. The eco-effects tests include acute, subacute, chronic, and field studies that are part of a tiered testing scheme. The results from one tier are evaluated to determine potential toxicological hazards, and if further testing is required in the higher tiers. The adverse effects examined include mortality, reduction in growth, reproductive impairment, changes in number of species, bioaccumulation of residues in nontarget organisms, and in higher tier studies, structure and function changes in the ecosystem. The data are used by the EPA to determine whether product labeling should carry warning statements pertaining to toxicity to birds, fish, or wildlife, and whether the product can be registered or should be subjected to special review.

Occupational and Residential Exposure.

Nontarget Insects and Nontarget Plants. Additional ecological hazard evaluation studies might be needed, using honey bees and certain plants including terrestrial and aquatic species (40 CFR 158.540 and 590, and Subdivision J and L Guidelines).

Incidents of illness in field workers, particularly in California, led to requirements for data to establish safe reentry intervals after treatment of fields with acutely toxic pesticides (40 CFR 158.390 and Subdivision K and U Guidelines). In addition to post-application data, such as foliar and soil residue dissipation, the EPA has also required development of monitoring data to estimate exposure in mixer loaders who handle pesticide concentrates, and in applicators who handle large quantities of the pesticide as diluted for use. Passive testing using absorbent patches fastened to clothing can lead to inaccurate estimation of exposure. Most exposure occurs on the hands of the workers. Data from many exposure studies conducted by individual companies, contract laboratories, academia, and government groups in the United States and Canada have been compiled on a generic basis to provide a better estimate of the amount likely to reach the skin of mixer loader applicators under actual working conditions. Studies have also been done on levels of certain pesticides or metabolites in blood or excreted in urine of exposed workers. These latter data cannot be interpreted without adequate information about metabolism and pharmacokinetics of the pesticide in humans (65).

Cost Estimates for U.S. Registration of Pesticides. Some environmental fate and some chronic toxicology studies cost more than U.S.\$1 × 10⁶ per study, and require as many as four years to complete, representing a large investment of both money and time for pesticide manufacturers. There is no assurance that registration can be obtained soon enough to recoup expenses amounting to up to \$70 × 10⁶ before the patent runs out (Fig. 2) (66). Furthermore, the EPA amended 40 CFR 152.404 in 1988 by proposing to assess fees for reviewing studies submitted in support of registration applications. Fees such as \$184,500 for a new chemical, \$64,000 for a new biochemical or microbial, \$33,000 for a new use pattern, \$4,500 for an experimental use permit, \$4,000 for an old chemical, or \$700 for an amendment to an existing application were proposed. However, the EPA has rescinded imposing these fees for applications submitted between October 25, 1988, and September 30, 1997, during which time registrants have to pay

increasingly large fees for review of studies submitted in support of reregistration of existing products.

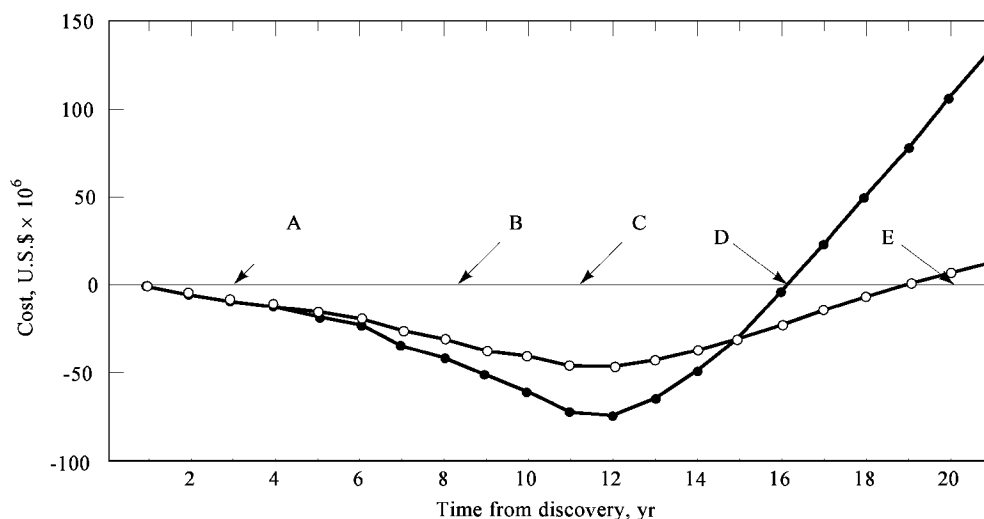


Fig. 2. (•) Cumulative and (°) 8% discounted cash flow rate for development of a hypothetical agrochemical in constant 1994 U.S. dollars, where A represents the time the patent was first issued; B, first registered use; C, manufacturing plant start-up; D, positive cumulative cash flow; and E, patent expiration (66). Courtesy of CRC Press.

Large fees are also assessed for review of studies submitted in support of a petition for residue tolerances in foods or feeds. In June 1994, 40 CFR 180.33 was amended to increase the fees to \$58,550 for the establishment of a new tolerance or a tolerance higher than already established, plus \$1,450 for each raw agricultural commodity (RAC) more than nine; and increased to \$13,400 for the establishment of a tolerance at the same or a lower numerical level or levels than a tolerance already established for the same pesticidal chemical, plus \$900 for each RAC. Tolerances or maximum residue levels in on RACs are regulated under provisions of Section 408 of the Federal Food Drug and Cosmetic Act, and tolerances for higher residues in processed fractions are regulated under Section 409. The latter includes the much debated Delaney Clause which forbids intentional or unintentional addition to food of substances that have been shown to cause cancer, regardless of the dose or significance of any risk anticipated from ingestion of the residue by humans. The Delaney Clause, passed by U.S. Congress in 1958, has as of this writing not been updated to take into account tremendous advances in analytical technology, resulting in the dichotomy of different regulatory standards for raw and processed foods (47,48).

Individual registrations are needed for each formulated product containing one or more approved AIs, and each product must bear a label approved for each use that is recommended on that label. The EPA charges an ever-increasing annual maintenance fee for each product label registration and each product must be registered in each state where it is sold and used, with appropriate fees charged by each state. States can also require a registrant to submit studies to be evaluated independently in that state, such as in California which sets its own standards for registration.

The EPA also requires the registration of any establishment that manufactures or formulates pesticides in the United States. Each establishment is required to maintain complete records of all production, whether for use in the United States or for export to other countries (40 CFR 167). Special labeling is required for export products that are not registered for use in the United States, and before the export can take place the exporter must receive and transmit to the EPA a signed purchaser agreement certificate acknowledging that the purchaser knows that specific product does not have a U.S. registration (40 CFR 168). These same products can be produced in other countries and exported more expediently. Thus, U.S. manufacturers are at a disadvantage in competing for international trade.

Reregistration of Existing Pesticides in the United States. In 1988, the U.S. Congress approved amendments to FIFRA that mandated acceleration of the reregistration program authorized in 1972. The EPA was to reevaluate the scientific database underlying some 600 cases representing about 1300 AIs contained in 45,000 pesticide products registered prior to November 1, 1984. A deadline of 1997 was given to complete the process. The concern was that some older agrochemicals might not have been tested adequately, and might be hazardous to human health or the environment. Any studies that did not meet these stringent standards and guidelines for registration of a new chemical would have to be repeated, submitted for review by scientists in the EPA's Office of Pesticide Programs (OPP), and accepted as meeting all requirements for new registration (28,31,67).

The status of the reregistration program in October 1994 has been summarized for AIs and products (68). The number of registered products declined by more than half between 1988 and 1994, because of the institution of annual maintenance fees. Registrants were reluctant to pay for some 19,000 products which were either no longer produced or had little or no use. A significant number of additional products were also voluntarily canceled by registrants owing to the high cost of conducting the required studies, particularly long-term toxicology and environmental fate studies on AIs, and of reregistration fees imposed by the EPA to review the study reports. This loss of many older products is of particular concern to farmers who raise minor crops, such as vegetables, fruits, nuts, herbs, ornamentals, trees, and turfgrass, on much smaller acreages than the principal crops such as wheat, soybeans, corn, and cotton (66,69).

In 1993, the U.S. General Accounting Office (GAO) estimated that EPA is not likely to complete the reregistration program until the year 2006 (70). The most significant factor delaying reregistration has been the quality of the submissions by registrants. Consequently, an intensive cooperative effort was undertaken involving scientists from the EPA, various other government agencies, and the pesticide industry, to identify and resolve underlying problems that most frequently cause studies to be rejected (71). Between 1992 and 1995, rejection rate analysis chapters were issued for five disciplines: residue chemistry studies in raw agricultural commodities and processed fractions used as food or feed, and metabolism studies in plants and animals; toxicology, including acute, subchronic, and chronic lifetime studies; environmental fate studies, including aerobic and anaerobic metabolism in soil and water, field dissipation studies, and photodegradation in air and on soil; ecological effects in mammals, birds, fish, and other aquatic organisms; and worker exposure studies. Some testing issues have been resolved and rejection rates have improved. Harmonization of requirements in the United States with those in other developed countries has not been achieved, however, so duplicative studies are often needed to satisfy minor differences in study protocols (31,62).

Role of International Organizations in Pesticide Regulation. Most developed countries have established laws and regulations that outline policies for the production, registration, and use of pesticides. Much as in the United States, these determine the risks and benefits associated with pesticides, and promote safe and effective use. The harmonization of regulatory standards has become of greater importance with the expansion of world agricultural trade and the movement of agricultural commodities among nations, particularly in compliance with the General Agreement on Tariffs and Trade (GATT) adopted in 1994.

The U.S. GAO has also reviewed pesticide standards and regulations among member countries of the expanded European Union and the Organization for Economic Cooperation and Development (OECD) (Table 5) (72). A high degree of uniformity exists among the surveyed nations, including the United States, with regard to the kinds of test data required to register food-use pesticides. However, similar data requirements do not necessarily mean that countries receive the same information about a pesticide product or evaluate it in a similar manner. For example, there is a divergence of scientific opinion concerning what regulatory approach is most appropriate for dealing with substances that show some oncogenic effects (tumors) only at very high, near-lethal doses as compared to those that cause cancer through a genotoxic mechanism (63). Also, most other countries do not require

analyses of commodities for pesticide metabolites and do not include metabolite residues in the expression of tolerance levels (56,73).

Table 5. Members of the European Union (EU)^a and the International Organization for Economic Cooperation and Development (OECD)^b

Country	EU	OECD	Country	EU	OECD
Australia	no	yes	Luxembourg	yes	yes
Austria	yes	yes	Mexico	no	yes
Belgium	yes	yes	the Netherlands	yes	yes
Canada	no	yes	New Zealand	no	yes
Denmark	yes	yes	Norway	no	yes
Finland	yes	yes	Portugal	yes	yes
France	yes	yes	Spain	yes	yes
Germany	yes	yes	Sweden	yes	yes
Greece	yes	yes	Switzerland	no	yes
Iceland	no	yes	Turkey	no	yes
Ireland	yes	yes	United Kingdom	yes	yes
Italy	yes	yes	United States	no	yes
Japan	no	yes	Yugoslavia	no	c

^a In 1994, the former European Economic Community (EEC or EC) became the European Union. It originated as the Common Market created by the Treaty of Paris in 1951 and the Treaty of Rome in 1958. Greece joined in 1981, Spain and Portugal in 1984, and Austria, Finland, and Sweden in 1994.

^b Mexico was accepted in 1994 as the 25th nation in the OECD.

^c In 1994 the former Yugoslavia had observer status.

There is strong support for harmonization of pesticide regulations among countries to avoid having to repeat expensive studies to meet each country's requirements. Steps toward this goal were marked by development of OECD's Guidelines for Testing of Chemicals (under revision as of this writing) and their Principles of Good Laboratory Practice updated in 1992 (74) and by issuance in 1991 of the European Council Directive 91/414/EEC known as the Registration Directive (62). The comprehensive nature of the Directive and its significance is that, like in the United States, the reregistration of all AIs is required within 10 years. Ultimately all products containing those AIs need to be reregistered. Although some 800 AIs are theoretically registered for use in one or more of the various OECD Member States, only about 350–400 are of any significance. By 1994, a list of the first 89 AIs to be reviewed had been issued. A deadline of April 30, 1995, had been given for submission of complete dossiers of information and for issuance of a second list of 90 additional priority AIs. In some cases, new data is expected to be needed to fill gaps which arise as a result of new data requirements imposed by the Directive (75).

Several other international agencies also take part in activities related to the safe use of pesticides, particularly in developing countries (76). The International Programme on Chemical Safety (IPCS) is a joint venture of the United Nations Environment Programme (UNEP), the World Health Organization (WHO), and the International Labour Organization (ILO). IPCS conducts and disseminates evaluations of how chemicals can influence the environment and human health. IPCS staff also develop different methods of assessing risk related to chemicals using data from laboratory, epidemiological, and related studies. Between 1976 and 1994, IPCS issued a series of 160 Environmental Health Criteria documents, 48 of which dealt specifically with one or more pesticides. The principles, concepts, and definitions used by panels of experts selected by the United Nations Food and Agriculture Organization (FAO) and the World Health Organization (WHO) to evaluate residue data and toxicology data, respectively, were published in 1990 (77). These expert panels hold an annual joint meeting on pesticide residues (JMPR) to review selected pesticides using data from all study reports provided by manufacturers and governments. The FAO panel recommends maximum residue limits (MRLs) for each pesticide in those commodities of importance in international trade. The WHO panel proposes an acceptable daily intake (ADI) for the pesticide, having an adequate margin of safety (generally 100-fold or more) over the lowest no-effect-level (NOEL) observed in the toxicology studies. Monographs summarizing all the data used in the evaluations are prepared and distributed to governments around the world. The recommended MRLs are reviewed by delegations from many countries during annual meetings of the Codex Committee on Pesticide Residues (CCPR). These MRLs are revised, if necessary, and are eventually accepted by the Codex Commission as standards for foodstuffs in international trade. This process can take several years, mainly because of different pesticide use patterns in various countries, depending on climate, the pests to be controlled, and the relative prices of available pesticides.

FAO has also developed a series of guidelines related to pesticide control that cover legislation, registration, efficacy data, post-registration surveillance, and environmental criteria (75). Its Working Party of Experts on the Official Control of Pesticides recommends specifications and methods of analysis for pesticide technical products and formulations to help prevent distribution and sale of illegal or fraudulent products in remote areas. In 1985, FAO adopted the International Code of Conduct on the Distribution and Use of Pesticides (78). FAO also shares operational responsibility with UNEP for overseeing the Prior Informed Consent (PIC) program, which was incorporated into the Code of Conduct in 1989. PIC is based on the principle that chemicals that are banned or severely restricted for health or environmental reasons should not be exported without the consent of relevant authorities in the recipient country. Although a long list of pesticides was initially proposed for PIC, an international panel reduced the number to those that are truly hazardous. However, the EPA's pesticide export policy has been strengthened (40 CFR 168.65–85) by requiring the exporter to notify the recipient if the product is not registered in the United States. Moreover, the EPA must receive a signed acknowledgment statement from the responsible authority in the recipient country before the product can be shipped.

Benefit and Risk Issues

As world population increases, urban expansion encroaches more and more on productive land areas used for growing food and fiber to feed and clothe all these people. Whereas crop yields may continue to increase with advances in agricultural technology, crop protection agents are needed to avoid losses owing to weeds, insects, and fungi (see *General References*). Pest infestations can affect the economies of entire regions. In the early to mid-1990s devastating insect-related cotton crop failures occurred in China, and in Pakistan and India (79). In the United States, cotton is making a comeback, because the boll weevil has been largely eradicated through the development of improved pesticides. Instead of spraying twice a week, each application costing \$7 to \$8 an acre, farmers might spray only twice a year.

Pesticides are subjected to extensive testing for residues in food, toxicology in laboratory animals, and fate in the environment before being registered for use. Moreover, uses are closely regulated by governmental agencies worldwide. Concerns exist about residues in food, especially for those pesticides classified as probably or possibly carcinogenic, based on studies in test animals given maximum tolerated doses over most or all of their respective lifetimes (63,64). As of mid-1995, legislation has been proposed but not acted upon to replace the 1958 zero-risk Delaney Clause with a more realistic regulation based on evaluations of no-significant risk levels.

Public concerns about pesticides in the diet of infants and children resulted in an expert committee convened by the U.S. National Academy of Sciences which devoted four years to the review of all available data. A consensus report was issued in 1993 (80). A number of recommendations for further work to more precisely define what constitutes the diet of infants and children were made. No risk could be estimated. The residue data reviewed by the panel were mainly from monitoring studies conducted by the FDA using multiresidue methods to analyze fresh produce and market basket samples collected from various geographic areas (81,82). These and other reliable scientific studies have demonstrated that relatively few food samples contain detectable residues. Most residues are far below established tolerances which are set above the maximum residue found in treated raw agricultural

commodities collected at the farm gate (83–85). No allowance for reduction of residues by trimming, washing, cooking, or other processing is made (86). The safety of actual residues in food has been supported by extensive risk assessment reviews conducted by the FoodSafe program at the University of California in Davis (84,85), and by the Council for Agricultural Science and Technology (CAST) which has issued brochures on pesticides food safety (87,88).

When illegal residues have been found in monitoring studies conducted by the FDA or USDA, the reason has often been that no U.S. tolerance had been requested for that particular pesticide in that specific crop. For example, an imported crop would be deemed to be adulterated and would be seized at the port of entry into the United States if found to contain a pesticide residue in the absence of a tolerance in that crop. This is so even if tolerances have been set for the same pesticide in several crops grown in the United States and the pesticide had been used to control a pest that does not exist in the United States. Furthermore, an international maximum residue level (MRL) might already have been established for that pesticide–crop combination under the Codex system of standards for food of importance in international trade. The U.S. GAO issued two reports on food safety and pesticides in 1991 (89,90).

Concern about pesticides contaminating surface or groundwater used for drinking purposes has resulted in groundwater monitoring (qv). Levels above health advisories (HAs) recommended by EPA have been found in samples taken from streams receiving runoff from freshly treated fields following heavy rainfall in the spring (91). This peak declined rapidly during the growing season. The rates of decline of herbicide residues in or on soil depend on the climate and on the chemical and biological nature of the soils being tested, as well as on the physical and chemical characteristics of individual pesticides (92,93).

BIBLIOGRAPHY

"Poisons, Economic" in *ECT* 1st ed., Vol. 10, pp. 871–877 by H. H. Shepard, U.S. Dept. of Agriculture; in *ECT* 2nd ed., Vol. 15, pp. 908–923, by R. L. Metcalf, University of California, Riverside; in *ECT* 3rd ed., Vol. 18, pp. 302–320 by M. R. Hadler, Sorex Limited. "Repellents" in *ECT* 3rd ed., Suppl. Vol., pp. 786–805, by D. A. Carlson, University of Florida.

1. R. Greenhalgh and T. R. Roberts, eds., *Pesticide Science and Biotechnology, Proceedings of the 6th IUPAC Congress of Pesticide Chemistry, Ottawa, 1986*, Blackwell Scientific Publications, Palo Alto, Calif., 1987, 604 pp.
2. H. Frehse, ed., *Pesticide Chemistry: Advances in International Research, Development and Legislation, Proceedings of the 7th International Congress of Pesticide Chemistry (IUPAC), Hamburg, 1990*, VCH Verlagsgesellschaft, Weinheim, Germany, 1991, 666 pp.
3. N. N. Ragsdale, P. C. Kearney, and J. R. Plimmer, eds., "*Options 2000*," *Proceedings of the 8th International Congress of Pesticide Chemistry, (IUPAC), Washington, D.C., 1994*, American Chemical Society, Washington, D.C., 1995, 450 pp.
4. D. R. Baker, J. G. Fenyes, W. K. Moberg, and B. Cross, eds., *Synthesis and Chemistry of Agrochemicals*, ACS Symposium Series No. 355, American Chemical Society, Washington, D.C., 1987.
5. D. R. Baker, J. G. Fenyes, and W. K. Moberg, eds., *Synthesis and Chemistry of Agrochemicals II*, ACS Symposium Series No. 443, American Chemical Society, Washington, D.C., 1991, 610 pp.
6. D. R. Baker, J. G. Fenyes, and J. J. Steffens, eds., *Synthesis and Chemistry of Agrochemicals III*, ACS Symposium Series No. 504, American Chemical Society, Washington, D.C., 1992, 504 pp.
7. D. R. Baker, J. G. Fenyes, and G. R. Basaran, eds., *Synthesis and Chemistry of Agrochemicals IV*, ACS Symposium Series No. 584, American Chemical Society, Washington, D.C., 1995, 650 pp.
8. J. B. Bourke, A. S. Felsot, T. J. Gilding, J. K. Jensen, and J. N. Seiber, *Pesticide Waste Management: Technology and Regulation*, ACS Symposium Series No. 510, American Chemical Society, Washington, D.C., 1992, 278 pp.
9. S. O. Duke, J. J. Menn, and J. R. Plimmer, eds., *Pest Control with Enhanced Environment Safety*, ACS Symposium Series No. 524, American Chemical Society, Washington, D.C., 1993.
10. P. A. Hedin, J. J. Menn, and R. M. Hollingworth, eds., *Natural and Engineered Pest Management Agents*, ACS Symposium Series No. 551, American Chemical Society, Washington, D.C., 1994, 550 pp.
11. K. D. Racke and J. R. Coats, eds., *Enhanced Biodegradation of Pesticides in the Environment*, ACS Symposium Series No. 426, American Chemical Society, Washington, D.C., 1990, 296 pp.
12. K. D. Racke and A. R. Leslie, eds., *Pesticides in Urban Environments: Fate and Significance*, ACS Symposium Series No. 522, American Chemical Society, Washington, D.C., 1993, 385 pp.
13. N. N. Ragsdale and R. E. Menzer, eds., *Carcinogenicity and Pesticides: Principles, Issues, and Relationships*, ACS Symposium Series No. 414, American Chemical Society, Washington, D.C., 1989, 246 pp.
14. N. N. Ragsdale and R. J. Kuhr, eds., *Pesticides: Minimizing the Risks*, ACS Symposium Series No. 336, American Chemical Society, Washington, D.C., 1987, 183 pp.
15. M. A. Saleh, J. N. Blancato, and C. H. Nauman, eds., *Biomarkers of Human Exposure to Pesticides*, ACS Symposium Series No. 542, American Chemical Society, Washington, D.C., 1994, 328 pp.
16. L. Somasundaram and J. R. Coats, eds., *Pesticide Transformation Products: Fate and Significance in the Environment*, ACS Symposium Series No. 459, American Chemical Society, Washington, D.C., 1991, 308 pp.
17. B. G. Tweedy, H. J. Dishburger, L. G. Ballantine, and J. McCarthy, eds., *Pesticide Residues and Food Safety: A Harvest of Viewpoints*, ACS Symposium Series No. 446, American Chemical Society, Washington, D.C., 1991, 360 pp.
18. J. M. Van Emon and R. O. Mumma, eds., *Immunochemical Methods for Environmental Analysis*, ACS Symposium Series No. 442, American Chemical Society, Washington, D.C., 1990, 239 pp.
19. W. Y. Garner, M. S. Barge, and J. P. Ussary, eds., *Good Laboratory Practice Standards: Applications for Field and Laboratory Studies*, ACS Professional Reference Book, American Chemical Society, Washington, D.C., 1992, 572 pp.
20. M. L. Leng, E. M. K. Leovey, and P. L. Zubkoff, eds., *Agrochemical Environmental Fate: State of the Art*, Lewis Publishers, CRC Press, Boca Raton, Fla., 1995, 410 pp.
21. G. J. Marco, R. M. Hollingworth, and J. R. Plimmer, eds., *Regulation of Agrochemicals: A Driving Force in Their Evolution*, American Chemical Society, Washington, D.C., 1991, 188 pp.
22. G. J. Marco, R. M. Hollingworth, and W. Durham, eds., *Silent Spring Revisited*, American Chemical Society, Washington, D.C., 1987, 196 pp.
23. J. O. Nelson, A. E. Karu, and R. B. Wong, eds., *Immunoassays of Agrochemicals: Emerging Technologies*, ACS Symposium Series No. 586, American Chemical Society, Washington, D.C., 1995, 376 pp.
24. F. R. Hall and J. W. Barry, eds., *Biorational Pest Control Agents: Formulation and Delivery*, ACS Symposium Series No. 595, American Chemical Society, Washington, D.C., 1995, 320 pp.
25. C. Tomlin, ed., *The Pesticide Manual*, 10th ed, incorporating the *Agrochemicals Handbook*, Royal Society of Chemistry, Cambridge, U.K., 1994, 1341 pp.
26. A. L. Aspelin, *Pesticide Industry Sales and Usage: 1992 and 1993 Market Estimates*, 733-K-94-001, U.S. Environmental Protection Agency, Washington, D.C., June 1994.
27. H. Geissbuhler, C. d'Hondt, E. Kunz, R. Nyfeler, and K. Pfister, in Ref. 1, pp. 3–14.
28. M. L. Leng, *CHEMTECH* 408–413 (July 1991).
29. W. Klassen, Ref. 3, pp. 20–32.
30. C. A. Edwards, in G. Ekstrom, ed., *World Directory of Pesticide Control Organizations*, 2nd ed., The British Crop Protection Council and the Royal

- Society of Chemistry, Crop Protection Publications, Cambridge, U.K., 1994, pp. 1–23.
31. B. Thomas, Ref. 21, pp. 73–79.
32. M. L. Leng, in Ref. 21, pp. 27–44.
33. NACA, *From Lab to Label: The Research Testing and Registration of Agricultural Chemicals*, National Agricultural Chemicals Association, Washington, D.C., Feb. 1993. NACA became the American Crop Protection Association in late 1994.
34. J. R. Finney, in Ref. 2, pp. 555–576.
35. Technical data, Wood Mackenzie Agrochemical Monitor No. 113, Agrochemical Company Service, County NatWest WoodMac Securities Ltd., London, Dec. 1994.
36. Technical data, Ref. Vol. Sect. 1, Agrochemical Company Service, County NatWest WoodMac Securities Ltd., London, Oct. 1994.
37. H. Iwamura and T. Fujita, in Ref. 1, pp. 61–64; E. L. Plummer, A. A. Liu, and K. A. Simmons, in Ref. 1, pp. 65–68; U. Muller, Ph. Huxley, and E. Ebert, in Ref. 1, pp. 69–72.
38. M. Los, in Ref. 1, pp. 35–42.
39. E. L. Plummer, in Ref. 2, pp. 51–60.
40. A. Dearing, in Ref. 2, pp. 61–72.
41. P. S. Magee, in Ref. 3, pp. 34–41.
42. S. V. Ley, in Ref. 2, pp. 97–107.
43. T. F. Tadros, in Ref. 3, pp. 76–86.
44. R. L. Denny, in Ref. 3, pp. 87–95.
45. R. L. Wilkins, in Ref. 3, pp. 96–103.
46. T. R. Shieh, in Ref. 3, pp. 104–114.
47. *Regulating Pesticides in Food: The Delaney Paradox*, Report of Board on Agriculture, Committee on Scientific and Regulatory Issues Underlying Pesticide Use Patterns and Agricultural Innovation, U.S. National Research Council, National Academy Press, Washington, D.C., 1987, 272 pp.
48. S. H. Abramson and C. S. Jablon, *EPA Interpretation of FFDCA Delaney Clause May Point the Way for Future Health and Safety Decisions*, Special Analysis, Chemical Regulation Reporter, Bureau of National Affairs, Inc., Washington, D.C., Apr. 19, 1991, pp. 51–56.
49. *Office of Pesticide Programs Annual Report for 1994*, EPA 735-R-95-001, U.S. Environmental Protection Agency, Washington, D.C., Jan. 1995, 59 pp.
50. *Pesticides: Reducing Exposure to Residues of Canceled Pesticides*, GAO RCED-95-23, U.S. General Accounting Office, Washington D.C., Dec. 1994, 42 pp.
51. *U.S. Code of Federal Regulations*, Title 40: Protection of the Environment, Subchapter E: FIFRA–Pesticide Programs, Parts 150 to 189, Rev. July 1, 1994, published as a Special Edition of the *Federal Register*, distributed by Government Institutes, Inc., Rockville, Md., 680 pp.
52. H. Frehse, in Ref. 1, pp. 293–300.
53. H. A. Moye, in Ref. 17, pp. 78–86.
54. B. D. Hammock and S. J. Gee, in Ref. 3, pp. 204–214.
55. J. N. Seiber, in Ref. 17, pp. 126–138.
56. D. B. Sharp, in Ref. 1, pp. 483–488; J. A. R. Bates, in Ref. 1, pp. 489–493.
57. C. L. Olinger, R. D. Schmitt, and E. Zager, in Ref. 2, pp. 513–521.
58. H. J. Nelson and S. C. Hermes, in Ref. 20, pp. 51–69.
59. P. L. Zubkoff, in Ref. 20, pp. 297–304.
60. D. S. Spatz, M. Shamin, S. Syslo, E. Regelman, and P. Caulkins, in Ref. 20, pp. 141–148.
61. H. M. Jacoby and E. M. K. Leovey, in Ref. 20, pp. 23–28.
62. D. J. Arnold, in Ref. 20, pp. 39–48.
63. W. L. Chen, E. W. Day, Jr., and J. E. Gibson, in Ref. 3, pp. 396–403.
64. A. Cockburn, in Ref. 3, pp. 405–417.
65. C. A. Franklin, in Ref. 2, pp. 411–422.
66. M. L. Leng, in Ref. 20, pp. 3–11.
67. E. C. Gray, in Ref. 21, pp. 45–54.
68. *Pesticide Reregistration Progress Report*, Office of Pesticide Programs, EPA 738-R-94-031, U.S. Environmental Protection Agency, Washington, D.C., Oct. 1994, 35 pp.
69. *Pesticides: Minor Uses/Major Issues*, Comments from CAST 1992-2, Council for Agricultural Science and Technology, Ames, Iowa, June 1992.
70. *Pesticides: Pesticide Reregistration May Not Be Completed Until 2006*, GAO RCED-93-94, U.S. General Accounting Office, Washington, D.C., May 1993, 37 pp.
71. *Pesticide Reregistration Rejection Rate Analysis: Summary Report*, EPA 738-S-95-001, U.S. Environmental Protection Agency, Washington, D.C., Feb. 1995, 66 pp.
72. *Pesticides: A Comparative Study of Industrialized Nations' Regulatory Systems*, GAO PEMD-93-17, U.S. General Accounting Office, Washington, D.C., July 1993, 104 pp.
73. J. A. R. Bates, in Ref. 2, pp. 493–503.
74. *Pesticides: Options to Achieve a Single Regulatory Standard*, GAO RCED-94-57, General Accounting Office, Washington, D.C., May 1994, 34 pp.
75. B. Thomas, in Ref. 3, pp. 382–388.
76. R. Lonngren, in G. Ekstrom, ed., *World Directory of Pesticide Control Organizations*, 2nd ed., The British Crop Protection Council and the Royal Society of Chemistry, Crop Protection Publications, Cambridge, U.K., 1994.
77. *Principles for the Toxicological Assessment of Pesticide Residues in Food*, IPCS Environmental Health Criteria Document No. 104, International Programme on Chemical Safety, Geneva, Switzerland, 1990, 117 pp.
78. *International Code of Conduct on the Distribution and Use of Pesticides*, United Nations Food and Agriculture Organization, New York, 1985; amended to include provisions for "Prior Informed Consent," 1990.
79. *Regal Comeback: King Cotton Reigns Once Again in South, with Demand and Production Both Up*, *Wall Street J.*, A1 and A8 (May 2, 1995).
80. *Pesticides in the Diet of Infants and Children*, Report of Board on Agriculture and Board on Environmental Studies and Toxicology, Commission on Life Sciences, U.S. National Research Council, National Academy Press, Washington, D.C., 1993, 386 pp.
81. M. Clower, Jr., in Ref. 17, pp. 105–113.
82. P. Lombardo and N. J. Yess, in Ref. 17, pp. 162–169.
83. J. F. McCarthy, in Ref. 17, pp. 182–191.
84. C. K. Winter, *Rev. Env. Contam. Tox.* **127**, 23 (1992).
85. C. K. Winter, *Regul. Tox. Pharmacol.* **15**, 137 (1992).
86. H. B. Chin, in Ref. 17, pp. 175–181.
87. *Pesticides and Safety of Fruits and Vegetables*, CAST No. 1990-1, Council for Agricultural Science and Technology, Ames, Iowa, Dec. 1990, 15 pp.
88. *Food Safety: The Interpretation of Risk*, CAST No. 1992-1, Council for Agricultural Science and Technology, Ames, Iowa, Apr. 1992.
89. *International Food Safety: Comparison of U.S. and Codex Pesticide Standards*, GAO PEMD-91-22, U.S. General Accounting Office, Washington, D.C., Aug. 1991.

90. *Pesticides: Limited Testing Finds Few Exported Unregistered Pesticide Violations on Imported Food*, GAO RCED-94-1, U.S. General Accounting Office, Washington, D.C., Oct. 1993, 34 pp.

91. D. A. Goolsby and W. A. Battaglin, in Ref. 20, pp. 159–173.

92. R. Grover, ed., *Environmental Chemistry of Herbicides*, Vol. I, CRC Press, Boca Raton, Fla., 1988.

93. R. Grover and A. J. Cessna, eds., *Environmental Chemistry of Herbicides*, Vol. II, CRC Press, Boca Raton, Fla., 1991.

General References

U.S. Federal Insecticide, Fungicide, and Rodenticide Act as Amended, FIFRA as passed in 1947; amended as the *Federal Environmental Pesticide Control Act* (1972), 540 09-89-012, Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, D.C., rev. Oct. 1988.

U.S. Code of Federal Regulations, Title 40: Protection of Environment, Subchapter E: FIFRA–Pesticide Programs, Parts 150–189, rev. July 1, 1994, published by the Office of the Federal Register as a Special Edition, distributed by Government Institutes, Inc., Rockville, Md., 680 pp.

Pesticide Assessment Guidelines, Data Requirements: Subdivision D, Product Chemistry; E, *Hazard Evaluation—Wildlife and Aquatic Organisms*; F, *Hazard Evaluation—Human and Domestic Animals*; G, *Product Performance*; H, *Labeling*; I, *Experimental Use Permits*; J, *Hazard Evaluation—Nontarget Plants*; K, *Exposure—Reentry Protection*; L, *Hazard Evaluation—Nontarget Insects*; M, *Biorational Pesticides*; N, *Chemistry—Environmental Fate*; O, *Residue Chemistry*; R, *Spray Drift Evaluation*; and U, *Applicator Exposure Monitoring*, Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, D.C., 1982 through current updates. Individual guidelines are available from the National Technical Information Service, Springfield, Va.

Catalog of OPP Publications and Other Information Media, EPA 730-B-94-011, Mar. 1994, 105 pp; available from Communications Branch, Office of Pesticide Programs, U.S. Environmental Protection Agency, Washington, D.C.

Drinking Water Health Advisories for Pesticides, Office of Drinking Water, U.S. Environmental Protection Agency, Lewis Publishers, Chelsea, Mich., 1989. Includes data used for evaluating 1-day, 10-day, and longer-term health advisories for 50 pesticides which have a potential for being found in drinking water, with specific references as sources of information.

F. M. Garfield, ed., *Quality Assurance Principles for Analytical Laboratories*, 2nd ed., AOAC International, Arlington, Va., 1991 (English), 1993 (Spanish). Includes forms used by U.S. federal agencies such as FDA and EPA.

P. A. Cuniff, ed., *Official Methods of Analysis of AOAC International*, 16th ed., Vols. I and II, AOAC International, Arlington, Va., 1995. Vol. I includes Pesticide Formulations and Pesticide Residues. Over 2100 collaboratively tested, approved methods for chemical and microbiological analyses, with 149 new methods, 103 revised updated methods, methods using antibody-based test kits, enzyme immunoassay, and annual supplements containing new and revised methods; chemical and common names of all drugs and pesticides; easy-to-locate references.

Pesticide Analytical Manual, Food and Drug Administration, Rockville, Md., Vol. I, 1990, and Vol. II, 1989. Includes multiresidue methods used by U.S. FDA to routinely check foods for pesticide residues, and single residue methods for certain pesticides not detected by any of the multiresidue methods.

T. Cairns and J. Sherma, eds., *Emerging Strategies for Pesticide Analysis*, CRC Press, Boca Raton, Fla., 1992, 368 pp. From the series *Modern Methods of Pesticide Analysis*; presents survey of improvements in extraction and cleanup, detailed descriptions of alternative analytical methods, latest applications, and up-to-date references.

T. Cairns and J. Sherma, eds., *Comprehensive Analytical Profiles of Important Pesticides*, CRC Press, Boca Raton, Fla., 1992, 304 pp. From the series *Modern Methods for Pesticide Analysis*; provides detailed information on properties and analytical methodology for nine prominent pesticides, pyrethroids, and fumigants in food. Includes formulations and uses, chemical and physical properties, toxicity data, and tolerances on various foods and feeds. Analytical information may be given in enough detail for methods to be carried out without having to consult additional literature sources.

K. A. Hassall, *The Biochemistry and Uses of Pesticides: Structure, Metabolism, Mode of Action and Uses in Crop Protection*, 2nd ed., VCH Verlagsgesellschaft, Weinheim, Germany, 1990.

S. Baker and C. Wilkinson, eds., *The Effects of Pesticides on Human Health*, Princeton Scientific Publishing Co., Princeton, N.J., 1990.

IPCS Health Criteria Document No. 104, International Programme on Chemical Safety, Geneva, Switzerland, 1990. Explains principles, concepts, and definitions used by the Joint FAO WHO Meeting on Pesticide Residues (JMPR) when assessing toxicological data on pesticide residues in food and establishing acceptable daily intakes. Guides, design and interpretation of relevant toxicological studies. Alerts readers to multiplicity of factors, from endocrinology of species to pharmacokinetic and pharmacodynamic properties of pesticide, that can influence validity of a study and determine relevance to safety assessment.

W. J. Hayes, Jr., and E. R. Laws, Jr., eds., *Handbook of Pesticide Toxicology*, Academic Press, Inc., San Diego, Calif., 1990. Three volume set; provides detailed toxicological profiles of more than 250 insecticides, herbicides, and fungicides; each compound described by identity, properties, and uses; toxicity to humans, laboratory animals, domestic animals, and wildlife; includes comprehensive coverage of diagnosis, treatment, prevention of injury, effects on domestic animals, wildlife, and humans; ~13,000 references.

Extension Toxicology Network, *EXTOXNET*, 2nd ed., available in hard copy and electronic form from Resource Center, Cornell University, Ithaca, N.Y., 1994. Contains 139 pesticide information profiles (PIPs), 16 toxicology information briefs (TIBS), and other information on current issues in pesticide toxicology and environmental chemistry.

Crop Protection Chemicals Reference (CPCR), 6th ed., Chemical and Pharmaceutical Press, John Wiley & Sons, Inc., New York, 1990. Includes brand names and product labels.

MSDS Reference for Crop Protection Chemicals, 1st ed, Chemical and Pharmaceutical Press, John Wiley & Sons, Inc., New York, 1989. Provides MSDS for crop protection chemicals found in CPCR, with updates every two months.

The Foundation for a Fantastic Tomorrow, Farm Chemicals Anniversary, 1894–1994, Meister Publishing Co., Willoughby, Ohio, Sept. 1994. Over 60 general articles on topics related to development and use of agrochemicals.

Farm Chemicals Handbook '95, Vol. 81, Meister Publishing Co., Willoughby, Ohio, 1995. Includes The Sine Index with cross referencing to alternative names for pesticide products, Fertilizer Dictionary, Pesticide Dictionary with chemical structures for active ingredients, Regulatory File with summary of various laws affecting production and use of agrochemicals, Environmental and Safety, Buyer's Guide, and Company Addresses.

C. Tomlin, ed., *The Pesticides Manual: A World Compendium, Incorporating the Agrochemicals Handbook*, 10th ed., The British Crop Protection Council and The Royal Society of Chemistry, Crop Protection Publications, Cambridge, U.K., 1994. Includes 725 entries by common name in alphabetic order, with chemical structure, chemical name(s), molecular formula, CAS Registry Number, physicochemical properties, commercialization, mode of action, uses, trade names, analytical methods, mammalian toxicology, ecotoxicology, and environmental fate.

G. Ekstrom, ed., *World Directory of Pesticide Control Organisations*, 2nd ed., The British Crop Protection council and the Royal Society of Chemistry, Crop Protection Publications, Cambridge, U.K., 1994. Includes introductory chapter by C. A. Edwards on *Pesticides as Environmental Pollutants*, and by R. Lungren on *International Efforts to Promote Chemical Safety*. A section by H. Kidd identifies international organizations and programmes operated under the United Nations and other intergovernmental organizations, and provides names and addresses of organizations responsible for control of pesticides in 159 countries.

P. H. Howard, ed., *Handbook of Environmental Fate and Exposure Data for Organic Chemicals*, Vol. III, *Pesticides*, Lewis Publishers, Chelsea, Mich., 1991. Provides chemical and physical properties of 70 pesticide active ingredients; reviews data on environmental fate and exposure potential, with general references.

"Risk Assessment of Pesticides," *Chem. Eng. News*, 27–55 (Jan. 7, 1991). Includes "Cancer Prevention Strategies Greatly Exaggerate Risks," by B. N. Ames and L. S. Gold, "Exposure to Certain Pesticides may Pose Real Carcinogenic Risk," by J. E. Huff and J. K. Haseman, "Insignificant Risks Must Be Balanced Against Great Benefits," by W. D. Carpenter, "Blundering Questions, Weak Answers Lead to Poor Pesticide Policies," by J. P. Myers and T. Colborn, and "Pesticide Regulation is Sound, Protective, and Steadily Improving," by D. D. Campt, J. V. Roelofs, and J. Richards.

E-C. Oerke, H-W. Dehne, F. Schnnbeck, and A. Weber, *Crop Production and Crop Protection—Estimated Losses in Major Food and Cash Crops*, Elsevier Science B. V., New York, 1994, 830 pp. Part 1 examines world population growth, agricultural production, and food supplies; Part 2 presents the development of

crop protection since the 1960s; Part 3 analyzes estimated losses in each of the world's major crops; Part 4 reflects on the changes to date and debates challenges faced, and likely direction to be taken by the crop protection industry.
D. T. Avery, *Saving the Planet with Pesticides and Plastic, The Environmental Triumph of High-Yield Farming*, Hudson Institute, Indianapolis, Ind., 1995, 432 pp.

Marguerite L. Leng
Leng Associates



PET AND LIVESTOCK FEEDS.

See FEEDS AND FEED ADDITIVES.

PETROLATUM WAXES.

See WAXES.

PETROLEUM

Nomenclature in the petroleum industry,
Origin of petroleum,
Composition,
Drilling fluids,
Enhanced oil recovery,
Refinery processes, survey,
Petroleum resources,

NOMENCLATURE IN THE PETROLEUM INDUSTRY

Crude oils, complex mixtures of naturally occurring organic liquids, are difficult to characterize in detail. Thus, many of the definitions used by the exploration, production, and refining sectors of the petroleum industry to describe petroleum and its products often lack precision. Even the term petroleum is poorly defined. Although often used synonymously with crude oil, petroleum is also frequently used to include natural gas (see GAS, NATURAL) and even solid hydrocarbons. Definitions of materials are commonly given in terms of the processes used to obtain them. Gasoline, for example, is the fraction of crude oil that distills between 15 and 200°C (60 and 392°F) (see GASOLINE AND OTHER MOTOR FUELS). Further complications arise because different parts of the petroleum industry use terms in differing ways. For example, wax may refer to material made up predominantly of long-chain alkanes, or it may refer to esters of long-chain alcohols and acids (see WAXES). Even the term hydrocarbons (qv) is used loosely indicating all the compounds in crude oils, whether or not these include compounds of nitrogen, sulfur, and oxygen.

In nature petroleum occurs in subsurface accumulations, or reservoirs, called fields that may be made up of one or more pools. Petroleum compositions vary widely and range from hydrocarbon-rich gases called natural gas, through crude oil liquids, to high molecular weight solids known as reservoir bitumen, residual oil, or tar. Petroleum is generated from kerogen, the high molecular weight, insoluble organic material in source rocks. High subsurface temperatures convert the kerogen to a petroleum-like range of compounds called bitumen. Part of this bitumen moves out of the source rock, in the process of expulsion or primary migration, and moves through permeable rocks to accumulate in a reservoir (secondary migration). The petroleum engineering procedures for bringing this petroleum to the surface are called production.

Traditionally the unit of crude oil production has been the barrel (bbl), equal to 42 U.S. gallons, 5.61 ft³, 158.8 L, or 0.159 m³. Increasingly petroleum reserves are given in metric tons, but because one unit is a volume and the other a weight, there can be no unique conversion factor for a material having a range of densities. Fields of >500 × 10⁶ bbl (79.5 × 10⁶ m³) of recoverable oil (>100 × 10⁶ bbl (15.9 × 10⁶ m³) in the U.S.) are called giants. Oil density may be reported in any appropriate units, and although metric units are used it is more common to report densities as degrees API (°API) or API gravity, where API stands for American Petroleum Institute. The relationship between density and API gravity is an inverse one defined by the following relationship:

°API = [141.5 / specific gravity at 60°F] - 131.5

Water corresponds to an API gravity of 10; crude oils fall between 10 and 60°API. The most common crude oil values are in the 35–40° range.

Other terms relating to physical properties include viscosity; refractive index; pour point, ie, the lowest temperature at which the oil flows; flash point, ie, the temperature at which the oil ignites; and aniline point, ie, the minimum temperature at which equal volumes of oil and aniline are completely miscible. These are determined under defined conditions established by ASTM.

Natural gas production is generally given in cubic feet or cubic meters (1000 ft³ = 1 Mcf = 28.3 m³). Reserves of a trillion cubic feet (Tcf) (28.3 × 10¹² m³) or more form a giant gas field. Natural gas is called dry when methane is the dominant hydrocarbon, and wet if it contains more than 4 L / 100 m³ of natural gas liquids (>0.3 gal / 1000 ft³). When gas (or oil) has a bad odor owing to high concentrations of hydrogen sulfide and volatile sulfur compounds it is called sour. Sweet gas has no noticeable odor. For statistical purposes gas is commonly reported as an equivalent amount of oil based on an equivalent heating capacity. The conversion is normally made using 170 m³ / 6000ft³ = 1 bbl and leads to a barrel of oil equivalent (boe).

Crude oils contain a wide range of hydrocarbons including straight and branched chains, ring compounds, and aromatics, as well as more complex compounds that incorporate nitrogen, sulfur, and oxygen (often called the NSOs), and some nickel and vanadium. The straight-chain, normal alkanes, range from 1 to >100 carbon atoms. These are often called paraffins in the petroleum industry because of the useful adjective paraffinic. Branched hydrocarbon chains that are nominally built up from repeated isoprene units (2-methyl butane structure) are called isoprenoids or terpenoids, and the 19- and 20-carbon compounds, named pristane and phytane, respectively, are frequently present in high concentrations. Isoprenoids also lead formally to saturated multiring structures. Petroleum chemists use the obsolete word naphthenes for the compounds that organic chemists call alicyclics. A better term, cycloparaffin, is used herein, leading to the adjective cycloparaffinic rather than naphthenic. Some of the characteristic structures in this group can be directly related to molecules synthesized by organisms. Whereas these have been called chemical fossils, it is more usual to call them biological markers or biomarkers. Common examples include the steranes and hopanes.

Aromatic hydrocarbons form a minor but important group of compounds in crude oils and range from single-ring to multiring compounds. The latter are called polycyclic aromatics (PAHs). Small aromatic molecules are environmentally significant and BTEX is commonly used as an abbreviation for benzene–toluene–ethyl benzene–xylenes. Multiringed compounds containing both aromatic and saturated rings may be referred to in the older literature as naphtheno-aromatics. The highest molecular weight fraction of crude oils commonly contains asphaltenes that are dark in color, NSO-rich, and very aromatic.

Most crude oil is refined to provide useful products and the dominant process is distillation (qv) (Table 1). Petroleum products produced by simple distillation without the use of pressure, cracking, or catalysts are called straight run. Residual material that has too high a molecular weight to distill forms a residuum, often called by such names as asphalt (qv). Naphtha (unrelated to naphthenes) is a distillate of petroleum having a boiling range lower than about 200 or 260°C (even occasionally up to 350°C). As a process intermediate, naphtha includes the components used to formulate gasoline and the

lighter grades of fuel oils such as kerosene and diesel fuel oil. As a finished product, naphtha usually denotes a more specific type of narrow boiling range material. The terms naphtha and solvent may be used interchangeably. For example, Varnish Makers' & Painters' (VM&P) naphtha has a range of 95–150°C. The majority of streams within a refinery designated as naphthas are straight-run materials, however the term can also be used for some cracked distillates.

Table 1. Generalized Distillation Ranges for Products Obtained During Crude Oil Refining

Product	Temperature range, °C	Carbon number range
gasoline	30–210	5–12
naphtha	100–200	8–12
kerosene and jet fuel	150–250	11–13
diesel and fuel oils	160–400	13–17
atmospheric gas oil	220–345	
heavy fuel oils	315–540	20–45
atmospheric residue	≥ 450	30 +
vacuum residue	≥ 615	60+

A number of other words that have traditionally been used in the petroleum industry are difficult to define precisely. These refer partly to specific boiling ranges, but also to certain intended uses. Thus, gasoline boils lower than naphtha, and kerosenes generally higher, but these terms are applied to products that are intended as fuels, rather than as solvents.

Gas oil is a product boiling slightly higher (235–425°C, or sometimes wider) than kerosene. The main feedstock to the catalytic cracking units (see FEEDSTOCKS), it received its name from use as an enriching agent in the production of city or manufactured gas. It is often used as diesel fuel.

Cylinder oil is a viscous oil used for lubricating the cylinders and valves of steam engines (see LUBRICATION AND LUBRICANTS). It is prepared from cylinder stock. The product from cylinder stock, when filtered and processed, is bright stock.

Cycle stock (recycle stock) denotes any product that is recycled, that is, taken back to an earlier stage in the process. The term cycle stock is also used for the gas oil-like product of catalytic cracking.

The word distillate is occasionally used by petroleum chemists with a specialized meaning. Although anything that has been distilled is, of course, a distillate, the term distillate is sometimes used to denote distillate fuel oil as opposed to residual fuel oil.

In the petroleum industry the International Union of Pure and Applied Chemistry (IUPAC) system is in widespread use for naming organic compounds. Two points, however, regarding group names and the prefix, iso, call for comment.

Group Names

Although the IUPAC system is effective in denoting any hydrocarbon, no matter how complicated, the system does not always result in convenient terms for groups of compounds. Because hydrocarbons having the same number of carbon atoms are apt to have boiling points within a small range, it would be convenient to have words that would refer to C₄, C₅, C , ... saturates, and C₄, C₅, C , ... monounsaturates, etc. The IUPAC system, however, goes by the number of carbon atoms in the longest straight chain. Thus, for example, the hydrocarbon referred to by the older systematic name of isobutane, when named in the IUPAC system is 2-methylpropane. However, for saturated aliphatic hydrocarbons, names such as butanes, pentanes, hexanes, etc, are taken as names in the older system, and therefore used as group names.

The situation is different when naming the ethylenic hydrocarbons, because the IUPAC has provided names such as propene, butene, and pentene, which are different from the former names ending in -ylene. However, butenes butylenes, pentenes pentylenes, etc, are not truly synonymous pairs, because the IUPAC name goes by the longest-chain rule. Isobutylene, named 2-methylpropene in the IUPAC system, would be included under substituted propenes, but not under butenes. Similarly the three pentylenes derived from isopentane are methylbutenes and not pentenes. For example, if it is necessary to denote the group of five isomeric monounsaturated hydrocarbons C₅H₁₀, the term pentylenes denotes this group, whereas pentenes denotes a narrower group having only two members, the two straight-chain pentylenes.

The Prefix Iso

In names such as isobutane, isopentane, isobutyl alcohol, and isoamyl alcohol, the prefix iso has a precise meaning, ie, one methyl group attached to the next-to-terminal carbon atom and no other branch. This notation is also frequently used by petroleum chemists to have a much wider meaning, denoting nothing more than branched-chain. If both meanings persist, any individual use of the prefix becomes ambiguous. Herein, an effort is being made to use branched-chain or just branched consistently for the looser meaning of iso, so that this prefix can be kept for denoting concisely what otherwise would require some circumlocution. An exception is made for the well-established name isooctane, which is 2,2,4-trimethylpentane [540-84-1].

BIBLIOGRAPHY

"Nomenclature in the Petroleum Industry" under "Petroleum" in ECT 3rd ed., Vol. 17, pp. 110–113, by C. Barker, University of Tulsa.

General References

N. J. Hyne, *Dictionary of Petroleum Exploration, Drilling, and Production*, PennWell Books, Tulsa, Okla., 1991, 625 p.
J. G. Speight, *The Chemistry and Technology of Petroleum*, 2nd ed., Marcel Dekker, Inc., New York, 1991, 760 p.
B. P. Tissot and D. H. Welte, *Petroleum Formation and Occurrence*, 2nd ed., Springer-Verlag, New York, 1984, 699 p.

Colin Barker
University of Tulsa

ORIGIN OF PETROLEUM

Petroleum is a naturally occurring complex mixture made up predominantly of carbon and hydrogen compounds, but also frequently containing significant amounts of nitrogen, sulfur, and oxygen together with smaller amounts of nickel, vanadium, and other elements. It may occur in solid, liquid, or gaseous form as asphalt (qv), crude oil, or natural gas (see *GAS, NATURAL*), respectively. The economic importance of crude oil and natural gas has stimulated considerable interest in their origin. Because of their fluid nature petroleum phases are mobile in the subsurface and may have accumulated far from the place where they formed. Herein, origin of petroleum means the combined processes that operate in nature and culminate in the petroleum that is present as commercial accumulations in permeable and porous reservoir rocks. The processes involved are petroleum generation, which occurs in the source rock; migration out of the source rock and to the reservoir; and maturation and alteration which operate to change the composition of the petroleum after it has accumulated in the reservoir.

A biogenic origin for the carbonaceous material in petroleum is widely but not universally accepted. An inorganic origin of petroleum has been proposed (1,2) and there is a dualist theory incorporating both biological and inorganic aspects (3). However, because inorganic processes generate racemic mixtures, the presence of optically active compounds in oils, especially the multiringed cycloalkanes (naphthenes), provides strong support for a biological hypothesis. Oils also contain the so-called chemical fossils or biomarkers, compounds having characteristic molecular structures that can be related to living systems. The compounds include isoprenoids, porphyrins, steranes, hopanes, and many others. The relative abundances of members of homologous series are often similar to those in living systems. The strong odd preference in the long-chain normal alkanes (>C_{20}) is particularly well documented (4). In addition, the lack of thermodynamic equilibrium among compounds (5), and the close association of petroleum with sedimentary rocks formed in an aqueous environment, suggests a low temperature origin. In this context, low temperature means less than a few hundred degrees Celsius as opposed to temperatures in the 700–1200°C range that characterize igneous processes involving silicate melts. The elemental composition of petroleum (C,H,N,S,O), the isotopic composition of oils, and the presence of petroleum-like materials in more recent sediments are consistent with a low temperature origin. The evidence supporting a biological source for the material that generates petroleum is extensive (6–8).

Organisms produce a wide range of organic compounds including significant amounts of biopolymers like proteins (qv), carbohydrates (qv), and lignins (see *LIGNIN*), together with a wide variety of lower molecular weight lipids (6–8). After the death of the organism, all or part of this organic material may accumulate in aquatic environments where the various compounds have very different stabilities. Some are metabolized in the water column by other organisms (including bacteria) and only the biochemically resistant material is incorporated into sediments. Survival of organic material depends on many factors but particularly the oxidizing or reducing nature of the system. Preservation is strongly favored in anoxic sediments (9). However, the formation of a petroleum accumulation requires more than just a concentration of the relatively low molecular weight hydrocarbons that are present in more recent sediments. Although $\text{C}_2\text{--C}_{10}$ hydrocarbons are present in extremely low (parts per billion (ppb) level) concentrations in organisms and sediments, these can account for up to 50% or more of the volume of some crude oils.

Compounds that are not synthesized by organisms are also reported in crude oils. Apparently these were formed from the available organic matter in what is thought to be the main process of petroleum generation. As the organic matter in sediments is buried in a reducing environment, and subjected to gradually increasing temperature and pressure, petroleum is generated as an intermediate in a transformation process that ultimately leads to methane and graphite (see *CARBON, NATURAL GRAPHITE*; *HYDROCARBONS*). Oxygen is first lost from the organic matter as carbon dioxide and water. Continued low temperature ($<^{\circ}\text{C}$) rearrangement takes place and the immature organic matter is converted to high molecular weight, insoluble material called kerogen. This becomes more aromatic and carbon-rich as it generates petroleum and evolves toward graphite. In contrast, the solvent-extractable lower molecular weight organic materials called bitumen or extractables increase in hydrogen content and progress through compositions typical of crude oils to those of gas (10).

The nature of organic material (kerogen) in source rocks controls whether oil or gas is generated, and also controls the composition of these products. In many cases, the type of organic matter can be recognized by microscopic examination after removing the mineral matrix with hydrofluoric and hydrochloric acids. Lignin-rich wood-derived materials appear to generate gas; surface coatings of plants that are rich in cutin and esters of long-chain acids and alcohols produce waxy oils that frequently have high pour points; organic matter derived from algae and other aquatic organisms seems to produce the normal range of crudes (6–8). Algal debris is frequently amorphous. The principle organic matter types can be identified because these produce distinct tracks across an atomic H/C versus atomic O/C diagram, the so-called van Krevelen diagram (11). With rising temperature, generally produced by increasing depth of burial, both the H/C and O/C atomic ratios decrease as the kerogen increases in carbon content. Much of the amorphous kerogen is hydrogen-rich and oxygen-poor, whereas the woody type is oxygen-rich and hydrogen-poor.

The relative amounts of the various types of organic matter is controlled by the environment where the sediments were deposited. For example, rivers transport lignin and surface coatings derived from land plants to the marine environment. As a consequence deltas are gas-prone environments (12) that frequently contain high pour-point crude oils rich in long-chain alkanes (13,14). Source rocks containing low percentages of organic matter generate insufficient oil to drive migration. The oil remains in the rock and is subsequently cracked to gas, so that poor source rocks only expel gas regardless of the organic matter type.

The ratio of bitumen to kerogen in sediments is initially low, but increases with depth of burial. It was shown in 1965 (15) that in the Los Angeles basin the increase becomes significant below about 3350 m, but in the adjacent Ventura basin the increase occurs deeper at about 4575 m. The geothermal gradients are different in these two basins, but the marked increase in bitumen:kerogen ratio occurs at the same temperature in both of them. Because the generation of bitumen from kerogen is the process of petroleum generation, this demonstrated clearly the important role of temperature. It also implied a minor (if any) role for pressure which differed by more than 50% between the two basins. Subsequent analyses of samples from many wells worldwide have shown similar trends. Wells have relatively low bitumen:kerogen ratios for the shallower section and rapidly increasing values deeper (7).

The petroleum generation process can be duplicated by laboratory pyrolysis. Higher temperatures are needed to produce these reactions in a few hours or days rather than the millions of years in nature (16,17). Both dry pyrolysis and hydrous pyrolysis have been used.

The generation of petroleum is nonbiological, induced by temperature, and influenced by available time. It follows essentially first-order kinetics (18) where an increase of 10°C roughly doubles the reaction rate at low temperatures. The low molecular weight compounds generated from the kerogen show none of the biological characteristics typical of compounds in more recent sediments. Therefore, at increasing depth, and hence increasing temperature, the bitumen fraction loses features such as odd-even predominance in long-chain normal alkanes, optical activity, and the predominance of four- and five-ringed cycloalkanes (11). These trends have been well-documented in many areas. The petroleum generation process can be treated quantitatively using models based on first-order kinetics (19). The Lopatin approximation (20) seems to give reasonable values for the onset of oil generation when properly calibrated (21), but more rigorous methods that treat generation as a series of parallel first-order reactions are more reliable (18,22–25).

An important exception to thermal generation is the bacterial formation of methane. The bacteria are anaerobic and effective in sulfate-free, anoxic conditions (26), and have long been recognized for their role in forming marsh gas. Bacterially produced methane is isotopically light, generally more negative than -55 parts per thousand (expressed as -55°‰) for carbon, relative to the Pee Dee Belemnite (PDB) standard, and contains only trace amounts of ethane and higher hydrocarbons. This methane may, however, have higher amounts of inorganic gases. Many of the huge natural gas

accumulations in Siberia, including the giant Urengoy field which has reserves of $5.9 \times 10^{12} \text{ m}^3$ (210 Tcf), are thought to be bacterially generated. These accumulations are usually shallow, are isotopically lighter than -59° (27), and have only a percent or less of higher hydrocarbons. Examples are also known from other parts of the world (28–30).

Most petroleum is found in reservoir rocks that have high permeabilities and porosities, where these properties have been developed by natural processes of sorting and winnowing that remove fine-grained particles, including organic materials. Reservoir rocks generally have insufficient organic matter to generate commercially significant quantities of petroleum (see OIL SHALE). It is believed that petroleum generation occurs in organic-rich source rocks, and that part of the bitumen then migrates to accumulate in reservoir rocks (6,7). This is the source rock concept. Clearly, migration has a critical role in linking the organic-rich source rocks to the reservoir. Solubilities of hydrocarbons in subsurface waters are generally too low to be significant in the petroleum migration process (31,32), and most recent studies have stressed the expulsion of a separate crude oil phase out of source rocks (33,34). It appears that source rocks develop high internal petroleum saturation caused by petroleum generation and the displacement of water during compaction (35,36). Upon continuing compaction and kerogen conversion, oil droplets are forced out of source rocks into adjacent permeable carrier beds. High pressures approaching the rock load (lithostatic) can develop, and induce near-vertical fractures that are important in providing migration pathways out of the source rock. In this case the pressure gradient that develops can overcome buoyancy, and oil may be expelled downward and out the bottom of source rocks, as well as out of the top. The alkanes are less strongly absorbed in the source rock and so are preferentially expelled. In contrast, the nitrogen–sulfur–oxygen compounds (NSOs) are most strongly adsorbed and thus depleted in the expelled oil (37). Migration efficiency varies widely and appears to be dependent on the organic matter content of the rock. High efficiencies are associated with high organic contents. The controlling factor is hydrocarbon (bitumen) saturation, and rocks having less than about 1.0 wt % organic matter do not generate sufficient bitumen. As a consequence no oil migrates from these rocks and they are not effective source rocks (18,38).

Buoyancy is the main driving force through the carrier beds and oils continue to move upward (toward shallower areas) until stopped at a slope reversal in a structural trap, or where permeability decreases as in a stratigraphic trap (39,40). Migration distances can be in excess of 100 km (35,39). Oil may be remobilized after its initial accumulation in the reservoir. Although in the simplest case this may involve only a simple relocation, it can lead to significant compositional changes if both gas and oil are involved. When an anticlinal reservoir is full to the spill point and has a gas cap over oil, any spilling off the bottom is oil, and the next shallower trap thus accumulates oil with no gas cap. This process of differential entrapment (41) leads eventually to oil in the up dip (shallower) reservoirs and gas in the deeper ones. Geological examples are given (42,43).

The composition of petroleum changes and evolves in the reservoir in response to changing conditions. Thermal maturation of crude oil is brought about by the increasing temperature that accompanies increasing depth of burial (44,45). Some large molecules are broken down into smaller fragments and the trend is for an increasing percentage of the lighter fractions as the oil progresses to lower densities in the sequence from oil, to lighter oil, to wet gas, and finally dry gas (46). The increasing hydrogen content implied by this sequence is provided by parallel reactions involving cyclization and aromatization. These residual molecules get steadily more aromatic and larger, and as the solvent properties of the oil change, these molecules are precipitated in the reservoir in a process called natural deasphalting (47). The solid precipitate also takes with it much of the NSOs, nickel, and vanadium, so that the producible oil is of better quality.

Large changes in petroleum composition can be produced by contact with flowing water (48,49). As the water moves past the oil in the reservoir, most water-soluble components are removed. These include the light ends, particularly the small aromatics, leaving a tar layer at the oil–water interface. If the water brings bacteria and oxygen into contact with the oils at temperatures below approximately 70°C, substantial changes in crude oil composition can result (49–51). The aerobic bacteria preferentially consume normal alkanes and may convert waxy crudes of high pour points to naphthenic crudes of low pour points. The removal of low density compounds leaves a heavier oil enriched in sulfur and nitrogen compounds because heavier fractions resist bacterial degradation and so accumulate in the residue. Water washing always accompanies biodegradation.

Biomarkers form a small percentage of bitumen and crude oils, but relative distributions and complex structures are modified by the various processes involved during petroleum generation and accumulation. These biomarkers are widely used for correlation studies, and for recognition and documentation of the progress of generation and maturation (52,53).

The overall petroleum system (54) that leads to the accumulation of oil and gas in natural reservoirs can be summarized as follows: organic matter is incorporated into sediments as these are deposited; possible shallow generation of biogenic methane; organic matter is converted to petroleum-like materials by the influence of increasing temperature with lower temperatures partially offset by longer times; part of the lower molecular weight material that is generated subsequently migrates from the source rock through permeable carrier beds to the reservoir; after the oil reaches the reservoir significant compositional changes may be produced by increasing temperature, water washing, and bacterial degradation.

BIBLIOGRAPHY

"Petroleum Origin" in *ECT* 1st ed., Vol. 10, pp. 97–109, by B. T. Brooks, Consultant; "Origin of Petroleum" under "Petroleum" in *ECT* 2nd ed., Vol. 14, pp. 838–845, by S. P. Ellison, Jr., University of Texas; in *ECT* 3rd ed., Vol. 17, pp. 113–118, by C. Barker, University of Tulsa.

1. V. B. Porfirev, *AAPG Bull.* **58**, 3–33 (1974).
2. T. Gold, in D. G. Howell, ed., *The Future of Energy Gases* U.S. Geological Survey Professional Paper 1570, Washington, D.C., 1993, pp. 57–80.
3. R. Robertson, *Nature*, **199**, 113–114 (1963).
4. E. E. Bray and E. D. Evans, *Geochimica et Cosmochimica Acta*, **22**, 2–15 (1961).
5. M. A. Bestougeff, in B. Nagy and U. Colombo, eds., *Fundamental Aspects of Petroleum Geochemistry*, Elsevier, Amsterdam, the Netherlands, 1967, pp. 77–108.
6. J. M. Hunt, *Petroleum Geochemistry and Geology*, W. H. Freeman, San Francisco, Calif., 1979, 617 pp.
7. B. P. Tissot and D. H. Welte, *Petroleum Formation and Occurrence*, Springer-Vedag, New York, 1984, 699 pp.
8. M. H. Engel and S. A. Macko, eds., *Organic Geochemistry, Principles and Applications*, Plenum Press, New York, 1993, 861 pp.
9. G. J. Demaison and G. T. Moore, *AAPG Bull.* **64**, 1179–1209 (1980).
10. C. Barker, *AAPG Continuing Education Course Note Series*, American Association of Petroleum Geologists, Tulsa, Okla., 1979, #10, 159 pp.
11. B. Tissot, B. Durand, J. Espitalie, and A. Combaz, *AAPG Bull.* **58**, 499–506 (1974).
12. M. T. Halbouty, R. E. King, H. D. Klemme, R. H. Dott, Sr., and A. A. Meyerhoff, *AAPG Memoir* **14**, 528–551 (1970).
13. B. D. Evamy and co-workers, *AAPG Bull.* **62**, 1–39 (1978).
14. H. D. Hedberg, *AAPG Bull.* **52**, 736–750 (1968).
15. G. T. Philippi, *Geochimica et Cosmochimica Acta*, **29**, 1021–1049 (1965).
16. M. D. Lewan, in Ref. 8, pp. 419–442.
17. C. Barker and L. Wang, *J. Anal. Appl. Pyrolysis* **13**, 9–61 (1988).
18. P. Ungerer, in M. L. Bordenave, ed., *Applied Petroleum Geochemistry*, Editions Technip, Paris, 1993, Chapt. 11.5, pp. 395–442.
19. P. Ungerer in B. Durand and F. Behar, eds., *Advances in Organic Geochemistry*, Pergamon Press, New York, 1990, pp. 1–26.
20. N. V. Lopatin, *Izv. Akad. Nauk SSSR, Seriya Geologicheskaya* **3**, 95–106 (1971).
21. D. W. Waples, *AAPG Bull.* **64**, 916–926 (1980).
22. A. K. Burnham and R. L. Braun, in Ref. 19, pp. 27–40.
23. T. M. Quigley, A. S. Mackenzie, and J. R. Gray, in B. Doligez, ed., *Migration of Hydrocarbons in Sedimentary Basins*, Technip, Paris, 1987, pp. 649–666.
24. B. P. Tissot, R. Pelet, and P. Ungerer, *AAPG Bull.* **71**, 1445–1466 (1987).
25. P. Ungerer, J. Burrus, B. Doligez, P. Y. Chenet, and F. Bessis, *AAPG Bull.* **74**, 309–335 (1990).

26. D. D. Rice and G. E. Claypool, *AAPG Bull.* **65**, 5–25 (1981).

27. V. I. Yermakov and co-workers, *Akad. Nauk. SSSR Dokl.* **190**, 196–199 (1970).

28. G. E. Claypool, C. N. Threlkeld, and L. B. Magoon, *AAPG Bull.* **64**, 1131–1139 (1980).

29. D. D. Rice and G. W. Shurt, *AAPG Bull.* **64**, 969–987 (1980).

30. D. D. Rice, in D. G. Howell, ed., *The Future of Energy Gases* U.S. Geological Survey Professional Paper 1570, Washington, D.C., 1993, pp. 583–606.

31. C. D. McAuliffe, *AAPG Continuing Education Course Note Series* American Association of Petroleum Geologists, 1978, #8.

32. L. D. Price, *AAPG Bull.* **60**, 213–244 (1976).

33. W. A. England and A. J. Fleet, *Petroleum Migration*, Special Publication No. 59, Geological Society, London, 1991, 280 pp.

34. B. Durand, *Organ. Geochem.* **13**, 445–459 (1988).

35. J. A. Momper, in Ref. 31.

36. P. Ungerer, E. Behar, and D. Discamps, in M. Bjoroy, ed., *Advances in Organic Geochemistry*, John Wiley & Sons, Inc., New York, 1981, pp. 129–135.

37. B. P. Tissot and R. Pelet, *Proceedings 8th World Petroleum Congress*, **2**, 35–46 (1971).

38. W. A. England, A. S. Mackenzie, D. M. Mann, and T. M. Quigley, *J. Geol. Soc., (London)* **144**, 327–347.

39. J. A. Momper and J. A. Williams, *AAPG Memoir* (35), 181–191 (1985).

40. T. T. Showalter, *AAPG Bull.* **63**, 723–760 (1979).

41. W. C. Gussow, *AAPG Bull.* **38**, 816–853 (1954).

42. D. Gill, *AAPG Bull.* **63**, 608–620 (1979).

43. B. H. Thomas, P. Moller-Pedersen, M. F. Whitaker, and N. D. Shaw, in B. M. Thomas and co-workers eds., *Petroleum Geochemistry in Exploration of the Norwegian Shelf*, Graham and Trotman, London, 1985, pp. 3–26.

44. C. W. D. Milner, M. A. Roger, and C. R. Evans, *J. Geochem. Explor.* **7**, 101–153 (1977).

45. W. L. Orr, *AAPG Bull.* **58**, 2295–2318 (1974).

46. C. Barker, in E. C. Donaldson, G. V. Chilingarian, and T. F. Yen, eds., *Enhanced Oil Recovery*, Elsevier, Amsterdam, the Netherlands, 1985, Chapt. 2, pp. 11–45.

47. A. J. Lomando, *AAPG Bull.* **76**, 1137–1152 (1992).

48. E. Lafargue and C. Barker, *AAPG Bull.* **72**, 263–276 (1988).

49. S. E. Palmer, in Ref. 8, pp. 511–533.

50. J. Connan, in J. Brooks and D. H. Welte, eds., *Advances in Petroleum Geochemistry*, Academic Press, London, 1984, pp. 300–330.

51. W. K. Seifert, J. M. Moldowan, and G. J. Demaison, *Organ. Geochem.* **6**, 633–643 (1984).

52. A. S. Mackenzie, in Ref. 50, pp. 115–214.

53. K. E. Peters and J. M. Moldowan, *The Biomarker Guide: Application of Molecular Fossils in Petroleum Exploration*, Prentice-Hall, Englewood Cliffs, N.J., 363 pp.

54. L. B. Magoon and W. G. Dow, *The Petroleum System—From Source to Trap*, AAPG Memoir 60, American Association of Petroleum Geologists, Tulsa, Okla., 1994, 644 pp.

Colin Barker
University of Tulsa

COMPOSITION

Petroleum, literally rock oil, describes a myriad of hydrocarbon-rich fluids present in source rocks and accumulated in subterranean reservoirs (see HYDROCARBONS). Petroleum can include three phases: gaseous (natural gas), liquid (crude oil), and solid or semisolid (bitumens, asphalt (qv), tars, and pitches) (1) (see COAL; GAS, NATURAL; TAR AND PITCH). The molecular composition of the liquid portion of petroleum contributing to the crude oil properties and behavior is discussed herein. Crude oils vary dramatically in color, odor, and flow properties. These properties often reflect the origin of the crude. Historically, physical properties such as boiling point, density (gravity), odor, and viscosity have been used to classify oils (2–4). Crude oils may be called light or heavy in reference to relative density (or specific gravity). Light crude oils are rich in low boiling and paraffinic hydrocarbons; heavy crude oils contain greater amounts of high boiling and asphalt-like molecules. The heavy oils tend to be more viscous, higher boiling, more aromatic, and contain larger amounts of heteroatoms. Likewise, odor is used to distinguish between sweet or low sulfur, and sour or high sulfur, crude oils.

Petroleum is thought to be derived from a variety of living organisms buried with sediments in previous geological eras. A small fraction of these organisms were trapped in oxygen-deficient (or reducing) environments where they escaped complete oxidation to carbon dioxide. Over tens or hundreds of millions of years, the residual organic material was subjected to a complex series of chemical changes known as diagenesis and catagenesis (2,3). In diagenesis, which occurs below 50°C, the organics undergo microbial action and some chemical reactions, resulting in dehydration, condensation, cyclization, and polymerization. During catagenesis, which occurs under a thermal stress of 50–200°C, the organics react with the surroundings by a combination of thermocatalytic cracking, decarboxylation, and hydrogen disproportionation to form petroleum in the sedimentary rocks. In the vast majority of cases the petroleum is not found where the precursors were laid down, but in reservoirs where accumulation occurs after migration from the source rocks through geologic strata (3,4) (see PETROLEUM, ORIGIN OF PETROLEUM).

The distribution of biomarker isomers, molecules that retain the basic carbon skeletons of biological compounds from living organisms, serves not only as a set of fingerprints for oil–oil and oil–source correlation (to relate the source and reservoir for exploration), but also to give geochemical information on organic source input (marine, lacustrine, or terrigenous source), age, maturity, depositional environment (clay or carbonate, oxygen levels, salinity, etc), and alteration (water washing, biodegradation, etc) (5,6).

Knowledge of the composition of petroleum allows the refiner to optimize conversion of raw petroleum into high value products (4). Originally, petroleum was distilled and sold as fractions, primarily for use in illumination and lubrication (see LUBRICATION AND LUBRICANTS). As of this writing (ca 1995), crude oil is sold in the form of gasoline (see GASOLINE AND OTHER MOTOR FUELS), solvents (see SOLVENTS, INDUSTRIAL), diesel and jet fuel (see AVIATION AND OTHER GAS TURBINE FUELS), heating oil, lubricant oils, and asphalts, or it is converted to petrochemical feedstocks (see FEEDSTOCKS, PETROCHEMICALS) such as ethylene (qv), propylene (qv), the butenes, butadiene, and isoprene. Modern refining uses a sophisticated combination of heat, catalyst, and hydrogen (qv). Conversion processes include coking, hydrocracking, and catalytic cracking to break large molecules into smaller fractions; hydrotreating to reduce heteroatoms and aromatics, thereby creating environmentally acceptable products; and isomerization and reforming to rearrange molecules to those having high value, eg, gasolines of high octane number.

A knowledge of the molecular composition of a petroleum also allows environmentalists to consider the biological impact of environmental exposure. Increasingly, petroleum is being produced in and transported from remote areas of the world to refineries located closer to markets. Although only a minuscule fraction of that oil is released into the environment, the sheer volume involved has the potential for environmental damage. Molecular composition can not only identify the sources of contamination but also aids in understanding the fate and effects of the potentially hazardous components (7).

Crude oils contain an extremely wide range of organic functionality and molecular size. The variety is so great that a complete compound-by-compound description for even a single crude oil is not likely. The composite molecular composition of petroleum can, however, be described in terms of three classes of compounds: saturates, aromatics, and compounds bearing the heteroatoms sulfur, oxygen, or nitrogen. Within each of these classes there are several families of related compounds. Table 1 lists some of the compounds typically found in petroleum crude oils. Structures are shown in Figure 1. The saturates include normal alkanes, branched alkanes, and cycloalkanes called paraffins, isoparaffins, and naphthenes, respectively (see PETROLEUM, NOMENCLATURE IN THE PETROLEUM INDUSTRY). Alkenes (olefins) are rare to the extent of being considered an oddity. Aromatics range from benzene to multiple fused-ring analogues such as naphthalene, phenanthrene, etc. Sulfur is found in both polar (thiol) and nonpolar (thioether and thiophenic) forms (8). The nitrogen and oxygen compounds are more likely to be found in polar compounds such as pyridines, pyrroles, phenols, carboxylates, amides, etc, than in nonpolar ones such as ethers (9). The distribution and characteristics of these molecular species account for the rich variety of crude oils.

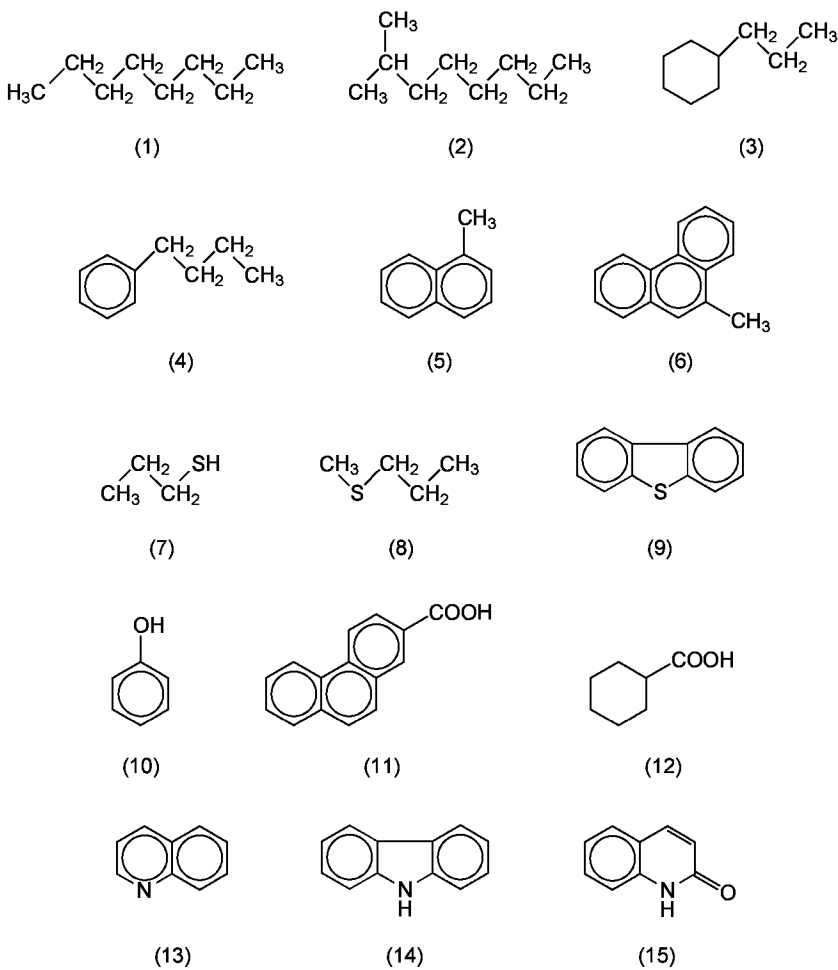


Fig. 1. Structures of compounds in petroleum crude oils. See Table 1.

Table 1. Compounds Found in Petroleum Crude Oils^a

Compound	CAS Registry Number	Molecular formula	Structure number
<i>Saturates</i>			
<i>n</i> -octane	[111-65-9]	C ₈ H ₁₈	(1)
2-methyloctane	[3221-61-2]	C ₉ H ₂₀	(2)
propylcyclohexane	[1678-92-8]	C ₉ H ₁₈	(3)
<i>Aromatics</i>			
<i>n</i> -butylbenzene	[104-51-8]	C ₁₀ H ₁₄	(4)
1-methylnaphthalene	[90-12-0]	C ₁₁ H ₁₀	(5)
9-methylphenanthrene	[883-20-5]	C ₁₅ H ₁₂	(6)
<i>Sulfur compounds</i>			
propyl mercaptan	[107-03-9]	C ₃ H ₈ S	(7)
methyl propyl sulfide	[3877-15-4]	C ₄ H ₁₀ S	(8)
dibenzothiophene	[132-65-0]	C ₁₂ H ₈ S	(9)
<i>Oxygen compounds</i>			
phenol	[108-95-2]	C ₆ H ₆ O	(10)
2-phenanthrene carboxylic acid	[40452-20-8]	C ₁₅ H ₁₀ O ₂	(11)
cyclohexyl carboxylic acid	[98-89-5]	C ₇ H ₁₂ O ₂	(12)
<i>Nitrogen compounds</i>			
quinoline	[91-22-5]	C ₉ H ₇ N	(13)
carbazole	[86-74-8]	C ₁₂ H ₉ N	(14)
2(1 <i>H</i>)-quinolinone	[59-31-4]	C ₉ H ₇ NO	(15)

^a See Fig. 1.

Elemental Composition

On an atomic basis, H :C ratios range from 1.5–2.0. The range of elemental composition of crude oil (3) may be given as follows:

Element	Composition range, wt %
carbon	84–87
hydrogen	11–14
sulfur	<0.1–8
oxygen	<0.1–1.8

nitrogen

<0.1–1.6

Nickel and vanadium are also generally present from a trace amount up to 1000 ppm. The value of using sulfur for petroleum classification can be clearly seen. Whereas the range of hydrogen and carbon are quite narrow, sulfur, the principal heteroatom in crude oil, varies significantly. The ranges for the other two heteroatoms (O and N) are also fairly wide, however in most crude oils these elements are nearly an order of magnitude lower than the sulfur level (Fig. 2). For example, the average level of nitrogen in >9000 crude oils is 0.094% , the average sulfur content is 0.65% (2). Although >90% of crude oils have nitrogen levels <0.2%, some crude oils have > 1.5% nitrogen. As noted in Figure 2b, the nitrogen concentrates into the highest boiling fractions. The trace metals Ni and V are generally orders of magnitude higher than other metals in petroleum, unless it is contaminated with coproduced brine salts (Na⁺, Mg²⁺, Ca²⁺, Cl⁻) or iron corrosion products gathered in transportation.

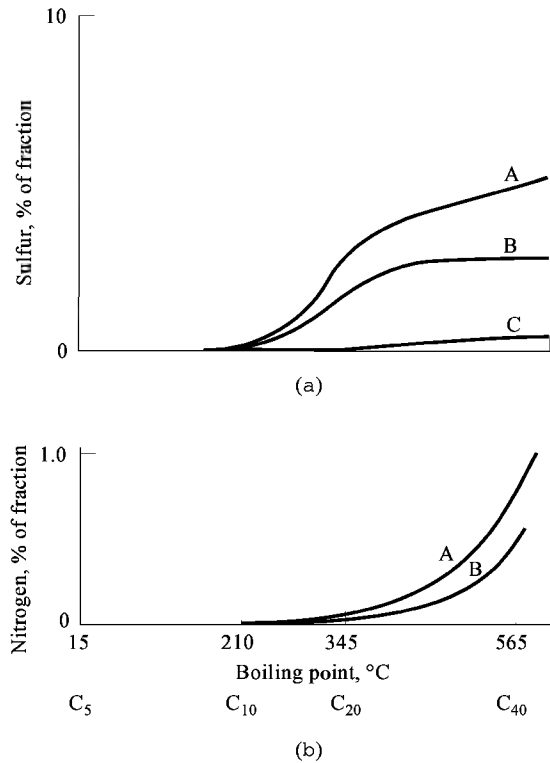


Fig. 2. Distribution of compounds as a function of crude oil boiling point: (a) sulfur where A, B, and C represent high (>2%), medium (ca 1.5%), and low (<0.1%) sulfur, respectively, and (b) nitrogen, where A and B represent high (ca 0.5%) and low (<0.1%) nitrogen, respectively.

Molecular Classes

The molecules in crude oil include several basic structural types (Table 1, Fig. 1). Because they may contain from 1 to 100 + carbon atoms and may occur in combination, the statistical potential for isomeric structures is staggering. For example, whereas there are just 75 possible paraffinic structures for C₁₀, there are >10⁵ isomers for C₂₀. A few structures tend to dominate the distributions of each isomer group, however.

The inclusion of naphthene and other aromatic rings introduces two additional dimensions, increasing the number of hydrocarbon isomers even further. A three-dimensional array in which the molecules could be described in terms of the number of aromatic rings, the number of naphthenic rings, and the number of carbons in alkyl side chains has been proposed (9). Conceptually, this amounts to describing a three-dimensional molecular mountain for hydrocarbons. There is also the potential of constructing similar mountains for heterocyclics. The two-dimensional image in the naphthenic and aromatic dimensions has been projected using sidebars to indicate the variation in alkyl substituents among crude oils (2).

Molecular characterization of a whole oil is beyond the capability of most analytical techniques. Distillation (qv), however, can separate petroleum into molecular weight fractions that simplify the task. Pioneering work with this approach, sponsored by the American Petroleum Institute (API) starting in 1925, has led to the identification of hundreds of individual compounds in distillation fractions of a single crude oil (10). More recently, developments in chromatography (qv) allowed oils to be fractionated by polarity as a second dimension. Under API sponsorship, the U.S. Bureau of Mines extended separations and measurement techniques to heavier fractions (11). At the same time, individual compounds have been isolated and quantified from increasingly higher boiling fractions (12–15). Techniques have been developed that use combinations of classical open-column adsorption chromatography, gel permeation chromatography, and ion-exchange (qv) separations to isolate fractions in which compounds could be identified by mass spectrometry (qv).

Whereas neither distillation nor chromatography achieves perfect separations among groups, the fractions generated are amenable to molecular characterization (14,15). An elegant argument has been made for the use of distillation for primary separation which is combined with a solubility step to achieve the atmospheric equivalent boiling point (AEBP) scale that covers room temperature to 1370°C (16). Because the distillations simplify analytical complexity, the bulk of available molecular compositional data on petroleum has been generated on sets of fractions defined by boiling points as gases (C₁–C₄), naphtha (initial bp – 210°C), mid-distillate (210–345°C), vacuum gas oil (345–565°C), and vacuum residuum (>565°C). The initial boiling point (ibp) of the naphtha depends on the amount of C₁–C₅ hydrocarbons dissolved in the oil. For characterization purposes, naphtha is usually de-pentanized so that the ibp is about 32°C. The material that boils above 345°C is referred to as atmospheric resid and is further distilled under vacuum into vacuum gas oil (VGO) and vacuum resid. Whereas the AEBP of the vacuum resid is >565°C, the actual temperature of distillation does not exceed 345°C to avoid thermal decomposition.

Crude distillations yield different quantities in each fraction. About the same amounts are distilled into the middle distillate and vacuum gas oil from conventional crude oils. More naphtha is distilled from light crude oils and more vacuum residuum is obtained from heavy crude oils (Fig. 3). The typical distribution of classes of petroleum compounds shows a significant shift with boiling point (Fig. 4). Whereas the lower boiling fractions are dominated by nonpolar saturated hydrocarbons that exist in limited isomeric forms, the higher boiling fractions increasingly contain a larger variety of classes, that have, in turn, an increasing number of possible isomers. As the boiling point increases, aromatic ring structures build in, first as naked rings, then more and more as rings having attached side-chain and naphthene ring carbons. Polar compounds, typically those having O and N functionality, that appear as trace impurities in the lower boiling fractions gradually become significant components in the higher boiling fractions. This is confirmed by the distribution of S and N in petroleum against boiling point (Fig. 2). The S, not including H₂S and the light sulfur compounds such as mercaptans and sulfides, present in petroleum gases, is more widely distributed than the nitrogen that concentrates in the highest boiling fraction. Not shown is the subtle decrease that occurs

in H/C ratio with increasing boiling point reflecting the increasing number of aromatic ring types at higher boiling point. The metals, nitrogen, and oxygen are predominantly found in the higher boiling fractions rich in polars.

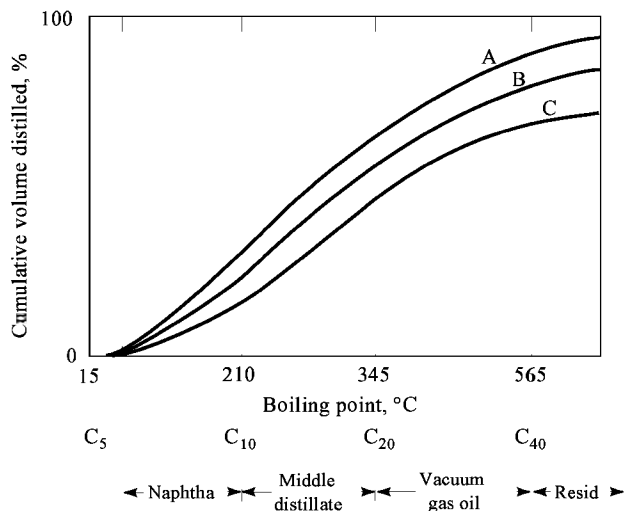


Fig. 3. Cumulative volume distilled as a function of boiling point from A, light; B, intermediate; and C, heavy crude oils (not including condensates).

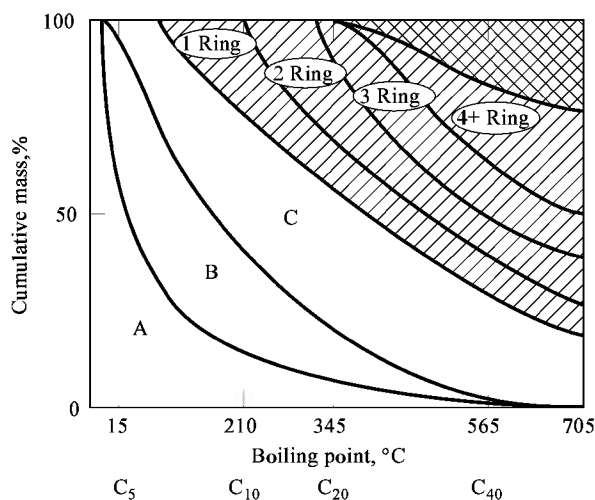


Fig. 4. Distribution of compound classes in crude oils as a function of boiling point. Region A represents normal paraffins; B, isoparaffins; C, naphthenes; □ the region of alkyl and naphthenic aromatics; and ▨ the region of polars.

Analytical Approaches. Different analytical techniques have been applied to each fraction to determine its molecular composition. As the molecular weight increases, complexity increasingly shifts the level of analytical detail from quantification of most individual species in the naphtha to average molecular descriptions in the vacuum residuum. For the naphtha, classical techniques allow the isolation and identification of individual compounds by physical properties. Gas chromatographic (gc) resolution allows almost every compound having less than eight carbon atoms to be measured separately. The combination of gc with mass spectrometry (gc–ms) can be used for quantitation purposes when compounds are not well-resolved by gc.

For the mid-distillates and vacuum gas oils (VGOs), class isolation and measurement techniques allow speciation of many compounds. In particular, multidimensional instrumental techniques have been applied. Capillary gc quantifies even closely related isomers. Element-selective detectors for gc are used to identify N, S, and O heterocompounds; gc–ms techniques are used to identify and quantify individual compounds or families of compounds. Characterization of families of compounds having similar degrees of unsaturation, expressed by π -number as in $C_nH_{2n+\pi}X$, where X represents heteroatoms, can be done by ms quantification (17). The combination of high performance liquid chromatography (hplc) and high resolution low voltage ms has also been used to resolve overlaps between aromatic hydrocarbons and sulfur analogues as well as among isomers of alkyl aromatics and naphthoaromatics (17). For VGOs ms techniques give semiquantitative results owing to the lack of reference compounds for calibrating relative sensitivities. Alternative techniques such as ultraviolet detection may be needed to distinguish among structural isomers (18,19).

The combination of chromatographic isolation followed by ms measurements has been extended well above 565°C into fractions of the vacuum resid (20–22). For fractions not amenable to that approach, techniques that provide average functionality data have been applied to describe petroleum composition (20). Established techniques such as titration for acidic and basic functionalities and for sulfur types, ultraviolet (uv) spectroscopy for aromaticity, and infrared (ir) spectroscopy for dipolar functionality have been supplemented with alternative instrumental techniques to expand the average database. These include nuclear magnetic resonance (nmr) for aromaticity, x-ray photoelectron spectroscopy (xps) for chemical bonding, and extended x-ray absorption fine structure (exafs) for atomic coordination environment, as well as improved Fourier-transform ir and Raman techniques for functionality (see INFRARED TECHNOLOGY AND RAMAN SPECTROSCOPY; MAGNETIC SPIN RESONANCE; SPECTROSCOPY). The individual average techniques do not provide molecular composition (23). On the other hand, these techniques can provide an average composition when used in combination. For those portions of the vacuum residua that are both nonvolatile and insoluble and hence not accessible to molecular speciation techniques, the alternative techniques are helpful.

Petroleum Gases and Naphtha. Methane is the main hydrocarbon component of petroleum gases. Lesser amounts of ethane, propane, butane, isobutane, and some C₄+ light hydrocarbons also exist. Other gases such as hydrogen, carbon dioxide, hydrogen sulfide, and carbonyl sulfide are also present.

The naphtha fraction is dominated by saturates having lesser amounts of mono- and diaromatics (Table 2, Fig. 4). Whereas naphtha (ibp to 210°C) covers the boiling range of gasoline, most raw petroleum naphtha molecules have a low octane number and most raw naphtha is processed further, to be combined with other process naphthas and additives to formulate commercial gasoline.

Table 2. Compounds Found in Petroleum Naphtha^a

Compound	CAS Registry	Molecular formula	Structure number
<i>n</i> -heptane	[142-82-5]	C ₇ H ₁₆	(16)
2-methylheptane	[592-27-8]	C ₈ H ₁₈	(17)
methylcyclopentane	[96-37-7]	C ₆ H ₁₂	(18)
ethylcyclohexane	[1678-91-7]	C ₈ H ₁₆	(19)
1,3-dimethylbenzene (<i>m</i> -xylene)	[108-38-3]	C ₈ H ₁₀	(20)
indan	[496-11-7]	C ₉ H ₁₀	(21)
naphthalene	[91-20-3]	C ₁₀ H ₈	(22)
tetrahydronaphthalene (tetralin)	[119-64-2]	C ₁₀ H ₁₂	(23)
decahydronaphthalene (decalin)	[91-17-8]	C ₁₀ H ₁₈	(24)
methyl mercaptan	[74-93-1]	CH ₄ S	(25)
3-methylthiacyclohexane	[5258-50-4]	C ₆ H ₁₂ S	(26)
3-(methylthio)pentane	[57093-84-2]	C ₆ H ₁₄ S	(27)

^a See Fig. 5.

Within the saturates in petroleum gases and naphtha, except for a few highly branched components in the C₈–C₁₀ range, every possible paraffin from methane to normal decane (*n*-C₁₀) is present. Depending on the source, one of the low boiling paraffins may be the most abundant compound in a crude oil, sometimes reaching a concentration of several percent. The isoparaffins begin at C₄. Isobutane is the only isomer of *n*-butane. The number of isomers grows rapidly with carbon number so that there are 74 isomers of C₁₀. Some of the individual 2-methyl isoparaffins may be present in concentrations >1%. Using instrumental techniques every possible isoparaffin in petroleum up to C₈ has been identified. Although complete resolution of all isomers >C₈ is beyond analytical capabilities, many of these compounds have also been identified.

In addition to the aliphatic (chain) molecules, the saturates contain cycloalkanes, called naphthenes, having mainly five or six carbons in the ring (Fig. 5). Methyl derivatives of cyclopentane and cyclohexane are commonly found in greater quantity than the parent unsubstituted structures and can be present at levels above 2% (2). Fused-ring dicycloalkanes such as decahydronaphthalenes (decalins) and hexahydroindans are also common, but nonfused bicyclic naphthenes, eg, cyclohexyl cyclohexane, are not.

The numerous aromatics in petroleum naphtha begin with benzene, but the C₁–C₃ alkylated derivatives of benzene generally are present in larger amounts. Toluene concentrations may reach nearly 2%; the combined xylene isomers exceed 1%; benzene, however, rarely exceeds 1% of a whole crude oil (2). Although present in lesser concentrations, each of the alkyl benzene homologues through the 20 isomeric C₄ alkyl benzenes have been isolated from Ponca City crude, as have several of the C₅-derivatives (24). Benzenes having fused cycloparaffin rings (naphthenoaromatics) such as indan and tetralin have been isolated along with a number of their methyl derivatives. Naphthalene is included in the naphtha, whereas the 1- and 2-methyl naphthalenes and higher homologues of fused two-ring aromatics appear in the mid-distillate fraction.

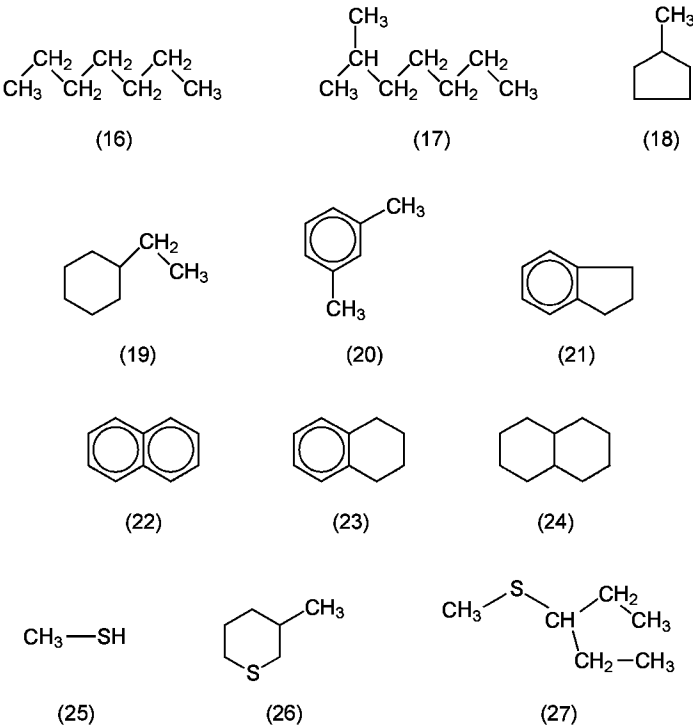


Fig. 5. Structures of compounds in petroleum naphtha. See Table 2.

Sulfur is the only heteroatom to be found in naphtha, and only at trace levels (Fig. 2). A systematic study of organic sulfur compounds in Wasson Texas crude, and to a lesser extent three other crude oils, culminated in the identification of over 200 compounds, most of which were in the naphtha fraction (24). However, the total of these compounds accounted for less than 1% of the sulfur in the whole crude. In sour petroleum naphthas, 50–70% of the sulfur can be found in the form of mercaptans, ie, thiols. Over 40 individual thiols have been identified, including all the isomeric C₁ to C₆ compounds plus some C₇ and C₈ isomers and thiophenol (25). In sweet naphthas, the sulfur is distributed between sulfides, ie, thio-ethers, and thiophenes. In these cases the sulfides may constitute more than 50% of the sulfur compounds in the form of both linear (alkyl sulfides) and five- or six-ring cyclic (thiacyclane) structures. Sulfur structure distribution mimics the hydrocarbons; ie, naphthenic oils having high cycloalkanes have high thiacyclane content, etc. Typical alkyl thiophenes in naphtha have multiple short side chains or exist as naphthenothiophenes (25). Though rare, methyl and ethyl disulfides have been confirmed to be present in some crude oils (26).

Mid-Distillates. As is indicated in Figure 4, saturates remain the primary component in the mid-distillate fraction of petroleum, but aromatics, which include simple compounds having up to three aromatic rings, and heterocyclics are present and represent a larger portion of the total. Some raw middle distillates are used directly as kerosenes, jet fuels, and diesel fuels; others are cracked and hydroprocessed before use. Some compounds found in middle distillates are listed in Table 3. Within the saturates, the concentration of *n*-paraffins generally decreases regularly from C₁₁ to C₂₀. The 2-methyl analogues are sufficiently unique to be seen as distinct peaks in gc analyses of mid-distillates. Few additional isoparaffins have been identified. However, the two isoprenoids (30) and (31) (see Fig. 6) are generally present in crude oils in sufficient concentration to be seen as irregular peaks alongside the *n*-C₁₇ and *n*-C₁₈ peaks in a gas chromatogram. These isoprenoids, believed to arise primarily as fragments of ancient chlorophyll [1406-65-1] have relevance, as simple biomarkers, to the genesis of petroleum. The distribution of pristane and phytane relative to the neighboring *n*-C₁₇ and *n*-C₁₈ peaks has been used to aid in the identification of crude oils and to detect the onset of biodegradation (3). The ratio of pristane to phytane has also been used for the assessment of the stage of oxidation reduction of the environment in which ancient organisms were converted into petroleum (5).

Table 3. Compounds Found in Petroleum Middle-Distillates^a

Compound	CAS Registry Number	Molecular formula	Structure number
<i>n</i> -hexadecane (cetane)	[544-76-3]	C ₁ H ₃₄	(28)
2-methylpentadecane	[1560-93-6]	C ₁ H ₃₄	(29)
pristane (2,6,10,14-tetramethylhexadecane)	[1921-70-6]	C ₁₉ H ₄₀	(30)
phytane (2,6,10,14-tetramethylhexadecane)	[638-36-8]	C ₂₀ H ₄₂	(31)
pentamethyldecalin	[80655-44-3]	C ₁₅ H ₂₈	(32)
fichtelite	[2221-95-6]	C ₁₉ H ₃₄	(33)
adamantane	[281-23-2]	C ₁₀ H ₁	(34)
biphenyl	[92-52-4]	C ₁₂ H ₁₀	(35)
fluorene	[86-73-7]	C ₁₃ H ₁₀	(36)
phenanthrene	[85-01-8]	C ₁₄ H ₁₀	(37)
2-methylbenzothiophene	[1195-14-8]	C ₉ H ₈ S	(38)
dibenzothiophene	[132-65-0]	C ₁₂ H ₈ S	(39)
benzothiacyclohexane	[2054-35-5]	C ₉ H ₁₀ S	(40)

^a See 6.

Mono- and di-cycloparaffins having five or six carbons per ring constitute the bulk of naphthenes in the middle-distillate boiling range, decreasing in concentration as the carbon number increases (2). The alkylated naphthenes appear to have a single long side chain as well as one or more methyl or ethyl groups (27). Similarly substituted three-ring naphthenes have been detected by gc ms (28). Generally, fused rings share just a single face; however, the highly symmetric multifused molecule adamantane and its alkyl-substituted homologues have been found (3,29). The most abundant aromatics in the mid-distillate are mono-, di-, and trimethyl naphthalenes. Other one- and two-ring aromatics are undoubtedly present in small quantities as either naphtheno or alkyl homologues in the C₁₁–C₂₀ range. In addition to these homologues of alkylbenzenes, tetralins, and naphthalenes, the mid-distillate contains some fluorenes and phenanthrenes, with traces of biphenyls (30). The phenanthrene structure is favored over that of anthracene structure (2). The *S*-heterocyclics in the mid-distillate range are primarily the thiacyclanes, benzothiophenes, and dibenzothiophenes. There are lesser amounts of dialkyl–, diaryl–, and aryl–alkyl sulfides (29). Alkylthiophenes are scarce or absent, but some evidence exists for benzthiacyclanes (Fig. 6). As for the naphtha fractions, these sulfur species account for a minimal fraction of the total sulfur in the crude.

[illegible]
$$\begin{array}{ccccccc}
 & \text{CH}_3 & & & & & \\
 & | & & & & & \\
 \text{CH}_3 - & \text{CH} - & \text{CH}_2 - & \text{CH}_2 - & \text{CH}_2 - & \text{CH}_2 - & \text{CH}_2 - & \text{CH}_2 - & \text{CH}_3 \\
 & | & & & & & & & \\
 & \text{CH}_2 & & & & & & &
 \end{array}$$

(29)

$$\begin{array}{ccccccc}
 \text{CH}_3 & & \text{CH}_3 & & \text{CH}_3 & & \text{CH}_3 \\
 | & & | & & | & & | \\
 \text{CH} & - \text{CH}_2 & - \text{CH} & - \text{CH}_2 & - \text{CH} & - \text{CH}_2 & - \text{CH} \\
 | & & | & & | & & | \\
 \text{CH}_3 & & \text{CH}_2 & & \text{CH}_2 & & \text{CH}_2 & & \text{CH}_3
 \end{array}$$

(30)

$$\begin{array}{ccccccc}
 & \text{CH}_3 & & \text{CH}_3 & & \text{CH}_3 & & \text{CH}_3 \\
 & | & & | & & | & & | \\
 \text{CH}_3 & -\text{CH} & -\text{CH}_2 & -\text{CH} & -\text{CH}_2 & -\text{CH} & -\text{CH}_2 & -\text{CH} & -\text{CH}_3 \\
 & | & & | & & | & & | \\
 & \text{CH}_2 & & \text{CH}_2 & & \text{CH}_2 & & \text{CH}_2
 \end{array}
 \quad (31)$$



 (32)



 (33)



 (34)

CC(C)C1CCC2C(C1)C(C)CCC2
(33)

(34)


(35)



(36)



(37)

Cc1cc2ccccc2s1
(38)c1ccc2c(c1)c3ccccc3s2
(39)c1ccccc1C2CCCCC2S
(40)

Fig. 6. Structures of compounds in petroleum middle-distillates. See Table 3.

Although only ppm levels of nitrogen are found in the mid-distillates, both neutral and basic nitrogen compounds have been isolated and identified in fractions boiling below 345°C (12). Pyrroles and indoles account for about two-thirds of the nitrogen. The remaining nitrogen is found in the basic pyridine and quinoline compounds. Most of these compounds are alkylated.

Vacuum Gas Oils

As is indicated in Figure 4, saturates contribute less to the vacuum gas oil (VGO) than the aromatics, but more than the polars present at percentage, rather than trace, levels. VGO itself is occasionally used as a heating oil but most commonly it is processed by catalytic cracking to produce naphtha or by extraction to yield lubricant oils.

Within the VGO saturates, distribution of paraffins, isoparaffins, and naphthenes is highly dependent on the petroleum source. The naphthenes account for roughly 60% of the saturates in a normal crude oil. However, samples can be found having paraffins from <20 to >80%. In most samples, the *n*-paraffins from C₂₀–C₄₄ are still present in sufficient quantity to be detected as distinct peaks in gc analyses. Some crude oils show a nearly symmetric pattern of peaks such that each carbon number is present in regular progression up to a maximum around C₂₇. Other crude oils show a similar distribution, but have preference for odd-numbered alkanes. Both the distribution and the selectivity toward odd-numbered hydrocarbons are considered to reflect differences in petrogenesis of the crude oils. Although *n*-paraffins are distinct in the gc, these usually account for only a few percent of the saturates measured by gc.

The bulk of VGO saturates consists of isoparaffins and especially naphthenes (Fig. 3). A few isoprenoid compounds, such as squalane, C₃₀; lycopane, C₄₀; and carotanes, C₄₀; have been detected. Analyses of petroleum waxes isolated from the saturates show a parallel lower level of 2- and 3-methyl alkanes as the most identifiable isoparaffins. Mass spectrometry techniques show that the naphthenes contain from one to more than six fused rings. Having an average carbon number of C₃₂, even the six-ring naphthenes have some alkyl substitution. For mono- and diaromatics, the alkyl substitution typically involves one long side chain and several short methyl and ethyl substituents. Some specific tetracyclic naphthenes, including steranes, and pentacyclic naphthenes, including hopanes, have been used as biomarkers (Fig. 7). These hopanes and steranes have also been used as nondegradable conserved internal markers for estimating biodegradation of crude oils during bioremediation processes (31).

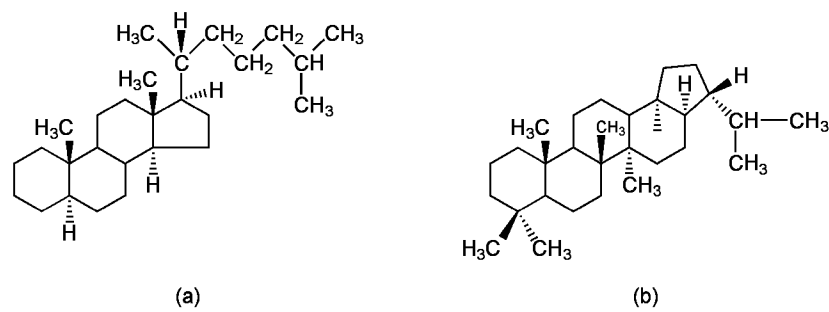


Fig. 7. Naphthenic biomarker compounds: (a) 5α(H),14α(H),17α(H),20(R)-cholestane [481-21-0], a C₂₇ sterane, and (b) 17α(H),21α(H)-hopane [13849-96-2], a C₃₀ pentacyclic triterpane.

The aromatics in VGO may contain one to six fused aromatic rings that may bear additional naphtheno rings and alkyl substituents in keeping with their boiling range. Mono- and diaromatics account for about 50% of the aromatics in petroleum VGO samples. Studies by nmr indicate an average of 3.5 substituents per monoaromatic, whereas ms techniques show the presence of up to four fused naphthenic rings on some aromatic compounds (19). This is consistent with the suggestion that these species originate from the aromatization of biogenic steroids (2,32). Although they are present at lower concentration, alkyl benzenes and naphthalenes commonly show one long side chain and multiple short side chains.

The fused 3 + ring aromatics in petroleum include both *cata*- and *peri*-condensed structures (see Table 4, Fig. 8). The *cata*-condensed species are those structures where only one face is shared between rings, the *peri*-condensed molecules are those that share more than one face. The fused ring aromatics form the class of compounds known as polynuclear aromatic hydrocarbons (PAH) which includes a number of recognized carcinogens in the 4 + ring family (33). Because of the potential health and environmental impact of PAH, these compounds have been studied extensively in petroleum.

Table 4. Fused-Ring Polynuclear Aromatic Hydrocarbons Found in Petroleum^a

Compound	CAS Registry Number	Molecular formula	Structure number
<i>cata</i> -Condensed aromatics			
phenanthrene	[85-01-8]	C ₁₄ H ₁₀	(37) ^b
chrysene	[218-01-9]	C ₁₈ H ₁₂	(41)
picene	[213-46-7]	C ₂₂ H ₁₄	(42)
<i>peri</i> -Condensed aromatics			
fluoranthene	[206-44-0]	C ₁₆ H ₁₀	(43)
pyrene	[129-00-0]	C ₁₆ H ₁₀	(44)
benzo[α]pyrene	[50-32-8]	C ₂₀ H ₁₂	(45)
benzoperylene	[11057-45-7]	C ₂₄ H ₁₄	(46)
coronene	[191-07-1]	C ₂₄ H ₁₂	(47)

^a See Fig. 8.
^b See Fig. 6.

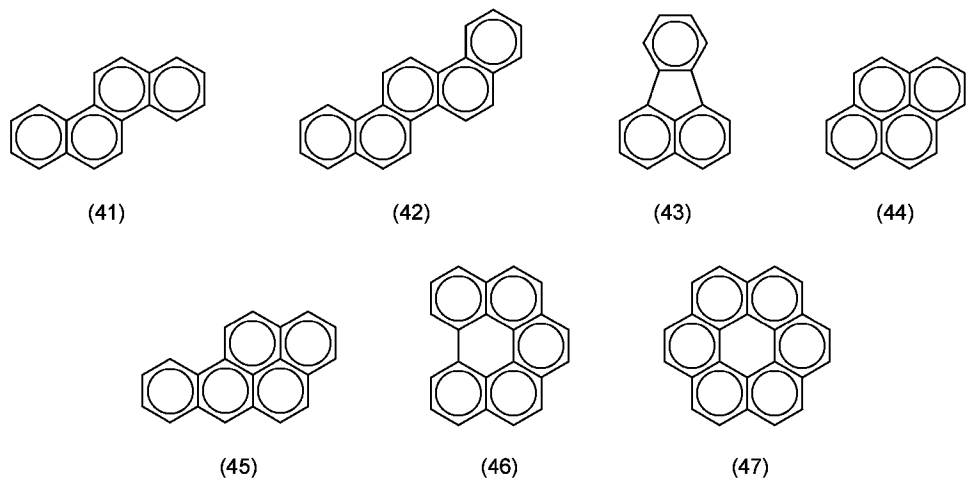


Fig. 8. Structures of fused-ring polynuclear aromatic hydrocarbons. See Table 4.

The total levels of three–six ring PAH in a petroleum VGO fraction range from 2 to 25%. However, the concentrations of individual isomers that have been quantified are generally expressed in parts per million (ppm). Although the most abundant reported individual phenanthrene compounds appear to be the C₁–C₃ derivatives, the average carbon number for the phenanthrenes in a VGO is C₃₂, indicating that the average side chain actually has 18 carbon atoms. The analytical techniques that have been used to isolate individual compounds are biased toward the least substituted aromatic structures, ie, the techniques that facilitate identification of a few target compounds eliminate the majority of multiring aromatic compounds.

Within petroleum certain aromatic structures appear to be favored. For example, alkyl phenanthrenes outnumber alkyl anthracenes by as much as 100:1. In addition, despite the bias in separation methods, alkyl derivatives appear to be more abundant than the parent ring compounds. For larger ring PAH, patterns exist, but are more difficult to detect owing to analytical resolution limits for the increasing number of isomers. A survey of crude oils by ms suggests that chrysenes are favored over pyrenes. Even for larger ring systems, the carbon numbers show that the parent compounds are missing or present in trace quantities. The C₃ derivatives are most prevalent.

The heterocyclics are significant contributors to the VGO fraction. Assuming an average molecular weight of 320 for the VGO, sulfur levels of

from 0.5 to >2.5% would indicate that from 5 to >25% of the molecules contain a sulfur atom (26). Results of API-60 show that 60° of the sulfur compounds are thiophenic and the remainder sulfidic. Although there is some contribution from alkylaromatic and diaryl sulfides, no dialkyl sulfides were detected. In contrast, from 20 to 36° of the sulfur compounds existed in thiacyclane structures of one to eight saturated rings. Gel permeation chromatography of the aromatic fractions coupled with ms led to identification of >30 homologous thiophenic series including 18 having more than a single sulfur atom. Although the distribution into these groups varied widely among crude oils, benzothiophenes, and dibenzothiophenes, having from zero to six naphtheno rings, were the prevalent thiophenic forms of sulfur, accounting for 30–40° of the total. Many of the homologous series appear to be *Δ*-analogues of the hydrocarbon PAH, ie, benzologues of dibenzothiophene, but no specific compounds have been isolated and identified unambiguously.

The nitrogen levels in crude oils are generally an order of magnitude lower than those of sulfur. In the VGO range, the nitrogen-containing compounds include higher molecular weight pyridines, quinolines, benzoquinolines, amides, indoles, and carbazoles; and molecules having two nitrogens, ie, diaza compounds, and three or four aromatic rings are especially prevalent (20). Typically, about one-third of the compounds are basic, ie, pyridine and its benzologues, whereas the remainder are present as neutral species such as amides and carbazoles. Although benzo- and dibenzoquinolines found in petroleum are rich in sterically hindered structures, hindered and unhindered structures have been found to be present at equivalent concentrations in source rocks. This has been rationalized as geo-chromatography in which the less polar (hindered) structures moved more readily to the reservoir (5).

Oxygen levels in the VGO parallel the nitrogen content. Thus, the most identified oxygen compounds are phenols and carboxylic acids, frequently called naphthenic acids. These may account for from ppm to nearly 3° of a VGO. The presence of numerous complex naphthenic and naphthenoaromatic acid structures in crude oils, especially immature forms, has been shown (34). Among the different structures a number of specific steroid carboxylic acids have been identified.

Vacuum Residua

The vacuum residua or vacuum bottoms is the most complex fraction. Vacuum residua are used as asphalt and coker feed. In the bottoms, few molecules are free of heteroatoms; molecular weights range from 400 to >2000, so high that characterization of individual species is virtually impossible. Separations by group type become blurred by the sheer mass of substitution around a core structure and by the presence of multiple functionalities in a single molecules. Simultaneously, the traditional gc and ms techniques require the very volatility that this fraction lacks.

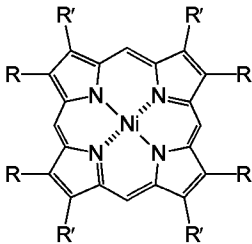
Classically, pentane or heptane precipitation is used as the initial step for the characterization of vacuum resid. The asphaltenes, the material that precipitates from the sample, range from 1 to 25° of vacuum resids. The soluble portion (maltenes) are then fractionated chromatographically into several poorly resolved classes for characterization. Asphaltene separations provide a simple way to remove some of the largest and most polar components; however, the fractions generated are still so complex that techniques such as nmr, titration, xps, or ir are mainly used to provide some average compositional detail. Asphaltenes are thought to be aggregates of complex molecular structures of MW <2000. The aggregation mechanism is not well understood. Evidence exists for both charge-transfer, among the numerous heteroatom functionalities, and ϕ – ϕ stacking of aromatics within the asphaltenes. The pitfalls of using average techniques for these complex systems have been discussed (21).

Additional general compositional information has been derived for the vacuum resid, using more sophisticated techniques that differ in approach toward isolating fraction. One approach uses functional group separations. Ion exchange (qv) has been used as the primary fractionation of vacuum resid into acids, bases, and neutrals (20,32). A second approach uses distillation into carefully defined atmospheric equivalent boiling points (AEBP) for the initial separation (21). Fractions generated by these separations have been subjected to a battery of techniques including liquid chromatography, field ionization mass spectrometry (fims), nmr, and elemental analyses.

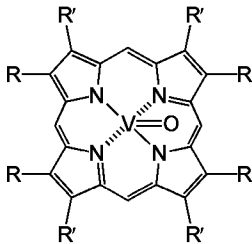
In fractions distilled above 565°C, the paraffins contribute only <2% of the 10–20° saturates. The one- to six-ring naphthenes contribute the rest, having average carbon numbers of 48–63. In the same study, aromatics shift progressively from an even distribution of mono- to tetraaromatics toward one dominated by five-ring types. As evidenced by the Σ -series data, many aromatics bear one or more naphtheno rings plus sufficient side-chain carbon atoms to bring the mean mol wt to nearly 800, ie, approximately 40 side-chain carbon atoms for a five-ring aromatic core of mol wt about 300. However, the elemental data show that 80° of the molecules could contain sulfur. Whereas most of the sulfur is thought to be thiophenic, studies by x-ray absorption spectroscopy have shown that in some vacuum resids it can exist in approximately equal amounts as sulfidic and thiophenic forms (35).

For the 565° + fraction the levels of nitrogen and oxygen may begin to approach the concentration of sulfur. These elements consistently concentrate in the most polar fractions to the extent that every molecule contains >2 heteroatoms. At this point, structural identification is not attempted. Rather, average characterization techniques are used to confirm the presence of functionalities such as acids, phenols, carbazole, and benzoquinoline, found in lower boiling fractions. Several models have been proposed based on the observed functionalities, apparent molecular weight, and elemental analysis of the fraction. These models suggest that molecules having boiling points greater than 565°C consist of multiple units similar to lower boiling components linked together with carbon and sulfur bridges rather than ever-increasing fused-ring structures (36).

The Ni and V concentrated into the vacuum resid appear to occur in two forms. From 10 to 14° of each of these two metals can be distilled in the 565–705°C boiling range, where they exhibit the strong visible Soret bands associated with the porphyrin structure. This tetrapyrrole structure (48,49), possibly derived from ancient chlorophyll, has been confirmed by a variety of analytical techniques.



(48)



(49)

Because the metalloporphyrins can provide insights into petroleum maturation processes, they have been studied extensively, and several families of related structures have been identified. On the other hand, the bulk of the metals are found in the heaviest fraction. Because those highly polar fractions do

not exhibit a Soret band, they are designated nonporphyrin metals. However, a number of techniques suggest that the metals in these Soret-inactive compounds still exist in molecules containing tetrapyrrole structures. Whereas other structures have been proposed (37), none has been demonstrated conclusively (38).

BIBLIOGRAPHY

"Petroleum Composition," in *ECT* 1st ed., Vol. 10, pp. 92–97, by A. N. Sachanen, Socony-Vacuum Oil Co., Inc.; in *ECT* 2nd ed., Vol. 14, pp. 845–855, by R. H. Hunt and M. J. O'Neal, Shell Oil Co.; "Petroleum–Composition," in *ECT* 3rd ed., Vol. 17, pp. 119–131, by J. J. Elliott and M. T. Melchior, Exxon Research and Engineering Co.

1. R. R. F. Kinghorn, *An Introduction to the Physics and Chemistry of Petroleum*, John Wiley & Sons, Inc., New York, 1983.
2. B. P. Tissot and D. H. Welte, *Petroleum Formation and Occurrence*, Springer-Verlag, New York, 1978.
3. J. M. Hunt, *Petroleum Geochemistry and Geology*, W. H. Freeman & Co., San Francisco, Calif., 1979.
4. J. G. Speight, *Fuel Science Technology Handbook*, Marcel Dekker, New York, 1990.
5. M. Yamamoto, *Adv. Org. Geochem.* **19**, 389 (1992).
6. K. E. Peters and J. M. Moldowan, *The Biomarker Guide: Interpreting Molecular Fossils in Petroleum and Ancient Sediments*, Prentice-Hall, Englewood Cliffs, N.J., 1993.
7. R. E. Jordan and J. R. Payne, *Fate and Weathering of Petroleum Spills in the Marine Environment*, Ann Arbor Science Publications, Ann Arbor, Mich., 1980.
8. M. F. Ali, M. Saleem, and S. M. Ghazali, *Hydrocarbon Process.* **60**(9), 129 (1981).
9. L. R. Snyder, *Accounts Chem. Res.* **3**, 290 (1970).
10. B. J. Mair, Z. Ronen, E. J. Eisenbraun and A. G. Horodysky, *Science* **154**, 1339–1341 (1966).
11. *Characterization of the Heavy Ends of Petroleum*, API Project 60, American Petroleum Institute, 1971; J. E. Dooley, D. E. Hirsch, C. J. Thompson, and C. C. Ward, *Hydrocarbon Process.* **53**, 187 (1974).
12. L. R. Snyder, *Anal. Chem.* **41**, 314 (1969).
13. L. R. Snyder, B. E. Buell, and H. E. Howard, *Anal. Chem.* **40**, 1303 (1968).
14. L. R. Snyder, *Anal. Chem.* **41**, 1084 (1969).
15. D. M. Jewell and co-workers, *Anal. Chem.* **44**, 1391 (1972).
16. M. M. Boduszynski, *Energy Fuels* **2**, 597 (1988).
17. C. S. Hsu, K. Qian, and Y. C. Chen, *Anal. Chim. Acta* **264**, 79 (1992); C. S. Hsu and co-workers, *Energy Fuels* **5**, 395 (1991).
18. P. Cleon, M. C. Foucheres, D. Cagniant, D. Severin, and W. Holstein, *Chromatographia* **18**, 543 (1985); J. M. Schmitter and co-workers, *Org. Geochem.* **6**, 579 (1984).
19. K. Qian and C. S. Hsu, *Anal. Chem.* **64**, 2327 (1992).
20. J. A. Green and co-workers, *Analysis of Heavy Oils: Method Development and Application to Cerro Negro Heavy Petroleum*, NIPER-452 (DE90000200, 2 vols.), IIT Research Institute, National Institute for Petroleum and Energy Research (NIPER), Bartlesville, Okla., 1989.
21. K. H. Algelt and M. M. Boduszynski, *Composition and Analyses of Heavy Petroleum Fractions*, Marcel Dekker, New York, 1994.
22. M. M. Boduszynski, *Liquid Fuel Technol.* **2**, 395 (1984).
23. C. S. Hsu and K. Qian, *Energy Fuels* **7**, 268 (1993).
24. B. J. Mair, *Oil/Gas J.*, 130 (Sept. 14, 1964).
25. H. T. Rall, C. J. Thompson, H. J. Coleman, and R. L. Hopkins, *Sulfur Compounds in Crude Oil*, Bulletin 659, U.S. Bureau of Mines, Washington, D.C., 1962.
26. V. S. Aksenov and V. F. Kamyranov, in R. Kh. Freidlina and A. E. Skorova, eds., *Organic Sulfur Chemistry*, Pergamon Press, New York, 1980, Chapt. 1.
27. R. J. Hood, R. J. Clerc and M. J. O'Neal, *J. Inst. Pet.* **45**, 168 (1959).
28. J. S. Richardson and D. Miller, *Anal. Chem.* **54**, 765 (1982).
29. St. Lanka and Sl. Hala, *Erdoel Kohle Erdgas Petrochem.* **11**, 698 (1958).
30. S. W. Lee, S. Coulombe, and B. Glavincevski, *Energy Fuels* **4**, 20 (1990).
31. R. C. Prince and co-workers, *Environ. Sci. Technol.* **28**, 142 (1994); C. S. Hsu, unpublished data.
32. J. F. McKay, J. H. Weber, and D. R. Latham, *Anal. Chem.* **48**, 891 (1976).
33. A. Dipple, in C. E. Searle, ed., *Chemical Carcinogens*, ACS Monograph 173, American Chemical Society, Washington, D.C., 1976, Chapt. 5, pp. 245–314.
34. W. K. Seifert and R. M. Teeter, *Anal. Chem.* **42**, 750 (1970).
35. G. N. George and M. L. Gorbaty, *J. Am. Chem. Soc.* **111**, 3182 (1989).
36. I. A. Wiehe, *Energy Fuels* **8**, 536 (1994).
37. R. H. Fish and J. Komlenic, *Anal. Chem.* **56**, 510 (1984).
38. C. D. Pearson and J. B. Green, *Energy Fuels* **7**, 338 (1993).

Winston K. Robbins
Chang Samuel Hsu
Exxon Research and Engineering Company

DRILLING FLUIDS

The success of any well-drilling operation depends on many factors, one of the most important of which is the drilling fluid. This fluid performs a variety of functions that influence the drilling rate, and the cost, efficiency, and safety of the operation (1–3). The drilling fluid or mud, as it is commonly called, is pumped down a hollow drill string through nozzles in the bit at the bottom of the well, and back up the annulus formed by the hole or casing and the drill string to the surface. The bit is turned by rotating the entire drill string from the surface or by using a downhole motor to only rotate the bit. After reaching the surface, the drilling fluid is passed through a series of vibrating screens, settling tanks or pits, hydrocyclones, and centrifuges to remove formation material brought to the surface. It is then treated with additives to obtain a set of desired physical and chemical properties. Once treated, the fluid is pumped back into the well and the cycle repeated.

Drilling fluids generally are composed of liquids, eg, water, petroleum oils, and other organic liquids; dissolved inorganic and organic additives; and suspended, finely divided solids of various types. The chemistry of the liquid phase and the level of suspended solids determine the treatment strategy and the efficiency of the mud-handling equipment. This chemistry also influences the type and amount of materials needed to maintain or change the density, viscosity, and other properties as required (4). Drilling fluids are dynamic systems. These muds constantly change in response to the changing conditions as the well is being drilled and downhole temperatures and pressures increase. Occasionally the downhole environment requires replacement of one fluid with another of a different type.

Drilling fluid costs range from several thousand to several million dollars per well and depend on the nature of the well being drilled. The length of the drilling time may vary from a few days to more than a year. On the average, about 6 to 8% of the total drilling cost arises directly from the drilling fluid and additives. Additional fluid and total well costs may arise from improperly formulated or treated fluids that can prolong the drilling time. Total worldwide sales of drilling fluid additives in 1994 was estimated to be \$1.2 × 10⁹. About 50% of this is spent in the United States.

The earliest fluid used for drilling oil wells was water. Various additives controlled the loss of water to permeable formations exposed in the borehole. As early as 1901, ground-formation material and surface clays (qv) were used to impart viscosity and density in the Spindletop well drilled near Beaumont, Texas. High density solids such as hematite and barite were applied in the early 1920s to increase the drilling fluid density, prevent influx of formation fluids, and control well blowouts. Drilling fluids were commercialized about 1926 (5).

Since 1980 over 1000 patents have been issued for drilling fluid systems and materials in the United States alone. A 1994 listing of products from 117 suppliers offers ca 3000 trade names (6). This array of trade name products actually represents less than 100 separate chemical types that may be purchased individually or as a blend. Moreover, some of these materials are for completion and workover fluids. These differ from drilling fluids in that completion fluids are used after the well has been drilled and prior to the initiation of production whereas workover fluids are used during remedial work on older wells.

Classification

Drilling fluids are classified as to the nature of the continuous phase: gas, water, oil, or synthetic. Within each classification are divisions based on composition or chemistry of the fluid or the dispersed phase.

Gas-Based Muds. Gas-based drilling fluids are used mostly for hard-rock drilling. These fluids range from compressed dry air or natural gas (see GAS, NATURAL) to water-based mist or stable foams (qv). Foam is considered gas-based. The gas is not the continuous phase, but it does comprise the bulk of the fluid volume. Bottomhole pressures imposed by a gas-based fluid are low, and therefore formation strengths must be relatively high, having little or no influx or formation fluid. Air, gas, or mist drilling requires a high annular velocity to remove drill cuttings. On the other hand, large cuttings can be removed at low annular velocities when using stable foams.

Chemical additives for gas-based drilling fluids are limited to surfactants (qv), certain polymers, and occasionally salts such as sodium or potassium chloride. An aqueous solution of the additives is injected into the air or gas flow to generate a mist or foam. No additives are used in dry air or gas drilling operations. Gas-based fluids are not recirculated and materials are added continuously. As the fluid exits the well, air or water vapor escapes to the atmosphere, gas and oil are burned, and water and formation solids are collected into a pit for later disposal. Stable foams must be destabilized to separate the air from the liquid phase for disposal.

Water-Based Muds. About 85% of all drilling fluids are water-based systems. The types depend on the composition of the water phase (pH, ionic content, etc), viscosity builders (clays or polymers), and rheological control agents (deflocculants or dispersants (qv)).

Fresh Water. Freshwater fluids can range from clear water having no additives to high density muds containing clays, barite, and various organic additives. Onshore wells typically use freshwater muds, as do some offshore wells where highly weighted muds are needed. Freshwater muds may be operated at pH levels ranging from 7 to 11. When drilling using clear water, small amounts of polymeric flocculants (qv) may be added to remove drill solids in a large settling pit in order to maintain a clean fluid for fast drilling.

When a viscous fluid is required, clays or water-soluble polymers (qv) are added. Fresh water is ideal for formulating stable drilling fluids as many mud additives are most effective in a system of low ionic strength. Inorganic or organic additives control the rheological behavior of the clays, particularly at elevated temperatures. An organic polymer may be used for filtration control. Mud pH is generally alkaline and, in fact, many viscosity control agents require an environment of pH > 9. Sodium hydroxide is by far the most widely used alkalinity control agent. Clay-base freshwater muds can be weighted to any desired density required to control formation pressures.

Seawater. Many offshore wells are drilled using a seawater system because of ready availability. Seawater muds generally are formulated and maintained in the same way that a freshwater mud is used. However, because of the presence of dissolved salts in seawater, more additives are needed to achieve the desired flow and filtration (qv) properties.

Salt Water. In many drilling areas both onshore and offshore, salt beds or salt domes are penetrated. Mud saturated with the salt present in the formation is used to reduce the hole enlargement that would result from salt dissolution by contact with an undersaturated liquid. In the United States, the salt formations are primarily made up of sodium chloride. In other areas, eg, northern Europe, the salt may be composed of mixed salts, predominantly magnesium and potassium chlorides. It has become quite common to use high (20–23 wt % NaCl) salt muds in wells being drilled in deep (>500-m water depth) water regions of the Gulf of Mexico. The reasons are twofold: stabilization of water-sensitive shales (7) and inhibition of the formation of gas hydrates (8).

The high salinity of salt water muds may require different clays and organic additives than those used in fresh- or seawater muds. Salt water clays and organic polymers contribute to viscosity. Filtration properties are adjusted using starch (qv) or cellulosic polymers. Alternatively, clays used primarily in fresh- and seawater muds can be prehydrated in fresh water and then added to a salt mud for viscosity and some filtration control. The pH ranges from that of the makeup brine, which may be somewhat acidic, to 9–11 through use of sodium hydroxide or lime. The presence of soluble calcium or magnesium in the mud determines whether it is feasible to maintain a high pH.

Calcium Treated. Fresh- or seawater muds may be treated with gypsum or lime to alleviate drilling problems that may arise from drilling water-sensitive shale or clay-bearing formations. Gyp muds (gypsum added) are generally maintained at a pH of 9–10, whereas lime muds (lime added) are in the 12–13 pH range. Calcium-treated muds generally require more additives to control flow and filtration properties than those without gypsum or lime.

Potassium Treated. Potassium treated systems combine one or more polymers and a potassium ion source, primarily potassium chloride, in order to prevent problems associated with drilling certain water-sensitive shales (9,10). The flow and filtration properties may be quite different from those of the other water-base fluids. Potassium muds have been applied in most active drilling regions around the world. Environmental regulations in the United States have limited the use of potassium muds in offshore drilling owing to the apparent toxicity of high potassium levels in the bioassay test required by discharge permits.

Low Solids/Nondispersed. Fresh water, clay, and polymers for viscosity enhancement and filtration control make up low solid nondispersed muds. Low solids muds are maintained using minimal amounts of clay and require removal of all but modest quantities of drill solids. These are called nondispersed systems because no additives are used to further disperse or deflocculate the viscosity building clays. Most water-based muds are considered dispersed because deflocculating additives are used to control the flow properties.

Nondispersed muds can be weighted to high densities, but are used primarily in the unweighted state. The main advantage of these systems is the high drilling rate that can be achieved because of the lower colloidal solids content. These are normally applied in hard formations where increasing the penetration rate can reduce drilling costs significantly and the tendency for solids buildup is minimal.

Oil-Based Muds. Oil-based drilling fluids have diesel or mineral oil (11) as a continuous phase with both internal water and solid phases. Fluids having no or very low water content are usually called oil-base muds or all oil muds; fluids having higher water contents are called invert oil-emulsion muds, or simply inverts. Most oil muds maintain a fixed oil-water ratio depending on the desired properties. Oil muds are employed for high angle wells where good lubricity is required, for high temperature wells where water-based systems may be thermally unstable, for drilling water-sensitive shale formations, or where corrosive gases such as hydrogen sulfide and carbon dioxide may be encountered. Environmental restrictions and cost often limit use, although higher drilling rates achievable using oil muds and polycrystalline diamond compact (PDC) bits can often offset the high fluid and disposal costs.

Oil-Base Muds. Oil-base muds have diesel or mineral oil as the continuous phase and either are formulated using no internal aqueous phase or have only a minimal water content (12). Organophilic clay or colloidal asphalt are used to control viscosity and filtration rates. The internal water phase, either added as part of the formulation or incorporated while drilling, is stabilized using emulsifiers. The desired density is reached by adding a powdered high specific gravity solid. A wetting agent ensures oil wetting of added or drilled formation solids.

Invert Oil-Emulsion Muds. Oil-based muds which incorporate an internal brine phase as an integral part of the formulation are called inverts or invert oil-emulsion muds. The water content can be increased up to 50 vol % of the total liquid phase using the proper emulsifiers and wetting agents. Suspension properties are achieved by the addition of organophilic clays. Colloidal solids such as oil-dispersible lignite (see LIGNITE AND BROWN COAL), resins, asphalt (qv), or gilsonite impart filtration control. Increased density is achieved using standard solid materials. The internal water phase is nearly always composed of a concentrated sodium or calcium chloride brine to provide a sufficiently reduced water activity to prevent osmotic uptake of drilling fluid water into water-sensitive formations.

Relaxed Fluid-Loss Oil Muds. Oil muds are usually characterized by very low filtration rates and at one time oil muds were not considered stable if other than minimal filtration rates were observed in standard tests. This feature contributed to the low drilling rates experienced for oil muds when using rock or natural diamond bits. Relaxing the normally tight filtration control specifications on oil muds can result in higher drilling rates without loss of emulsion or mud stability (13). These relaxed fluid-loss muds also are termed low colloid oil muds or relaxed filtration oil muds because the higher filtration rates are achieved by omitting some of the colloidal solids from the formulation and reducing the concentration of emulsifiers and surfactants in the fluid. Most relaxed fluid-loss oil muds contain 5–15 vol % brine, depending on the mud density. In most other respects these fluids are similar to a conventional invert oil-emulsion mud and can easily be converted to an invert. Use of relaxed filtration oil muds has dropped considerably owing to improvements in bit design. These improvements allow high drilling rates using the more conventional low filtration rate oil muds.

Synthetic-Based Muds. A new class of drilling muds, the synthetic-based muds, has been introduced to counteract the high costs associated with disposal of drill cuttings generated when diesel or mineral oil-based muds are used. These newer fluids, similar in formulation and performance to oil-based muds, have a continuous phase that consists of a synthetic organic liquid (14). Because of the similarity, synthetic-based muds are often called pseudo-oil muds outside the United States. A variety of fluids have been used as the continuous synthetic phase including an ester of a vegetable oil fatty acid (15,16), a polyalphaolefin (17,18), an acetal (19), linear alkyl benzenes, a low viscosity nonpetroleum hydrocarbon (20), etc. As of this writing (ca 1995) synthetic-base fluids are being developed and field tested at a rapid rate.

The synthetic-based fluids are much more expensive than mineral oil systems, but drill cuttings from offshore locations can often be discharged on site depending on environmental regulations. In some drilling areas, biodegradation rate is paramount; in others, such as in U.S. waters, toxicity and static sheen are key. The high cost of using synthetic fluids is offset by high performance for fast drilling and on-site discharge of cuttings. Both are cost saving operations.

Properties

The drilling fluid removes formation cuttings from beneath the bit and transports these cuttings to the surface, prevents influx of formation fluids into the wellbore, seals exposed permeable formations, maintains the stability of exposed formations, cools and lubricates the bit and drill string at points of contact with the cased or uncased borehole, and helps to suspend the weight of the drill string and casing. Furthermore, the fluid should not damage productive formations, it should not be corrosive to the equipment, and it must be cost effective. The fluid must be safe for handling and be compatible with the environment or be disposable in an environmentally sound manner.

How well the drilling fluid fulfills these functions is determined primarily by the response of the well. The whole drilling operation suffers if the fluid is not adequate. In extreme cases, drilling may be stopped or a hole may have to be redrilled. A variety of physical and chemical properties are monitored to ensure satisfactory performance and guide fluid formulation and treatment (21,22).

Density. The density of the drilling fluid is adjusted using powdered high density solids or dissolved salts to provide a hydrostatic pressure against exposed formations in excess of the pressure of the formation fluids. In addition, the hydrostatic pressure of the mud column prevents collapse of weak formations into the borehole. Fluid densities may range from that of air to >2500 kg m⁻³ (20.8 lb gal). Most drilling fluids have densities >1000 kg m⁻³ (8.33 lb gal), the density of water. The hydrostatic pressure imposed by a column of drilling fluid is expressed as follows:

P = 0.098 L ρ_m (0.052 L ρ_m)

where P = the hydrostatic pressure in kPa (psi); ρ_m = the drilling fluid density, kg m⁻³ (lb gal); and L = the column length or well depth, m (ft). Wells with bottomhole pressures of 100 MPa (14,500 psi) at 5000 m (16,400 ft) are not unusual. Such a well would require a mud density >2040 kg m⁻³ (17 lb gal) to prevent formation-fluid influx. At least 2 MPa (290 psi) above formation pressures normally is recommended to ensure well control. Bottomhole pressures imposed by the mud column that are higher than necessary can reduce drilling rates and induce fractures in exposed formations allowing loss of drilling fluid to the fractures (lost circulation). Operational safety can be adversely affected if lost circulation occurs.

Flow Properties. The fluid viscosity and annular flow velocity must be high enough to remove cuttings generated by the drill bit and other formation material that may fall into the wellbore. These solids are carried up the annulus to the surface where they are separated with varying degrees of efficiency and then disposal is effected. In order to accomplish this, low viscosity drilling fluids are circulated at high flow rates or high viscosity fluids at low flow rates. In addition, for maximum drilling rate, a low effective viscosity is desired at shear rates generated through the bit nozzles (10,000–100,000 s).

The varying demands on the flow properties are best met by fluids exhibiting non-Newtonian rheological characteristics. Drilling fluids are normally shear thinning, having apparent viscosity decreasing with increasing shear rate. Thixotropic properties characterized by a gel strength provide suspension of solids when mud circulation stops. The rheological properties are adjusted using inorganic and organic additives to achieve the capacity to carry cuttings and generate adequate bit hydraulic horsepower for fast drilling at moderate flow rates. The high flow rates required to remove drill cuttings using low viscosity muds and in high angle wells can increase pump maintenance and induce fracturing and lost circulation because of excessive bottomhole pressures (see also RHEOLOGICAL MEASUREMENTS).

Filtration Properties. Drilling fluids have a natural tendency to flow into permeable formations because the borehole pressure is generally

higher than that in the formation. To prevent excessive leak-off, a thin, low permeability filter cake is formed using additives. Filtration (qv) occurs under both dynamic (during circulation) and static (no circulation) conditions. Additives may affect each of these filtration conditions differently. The filtration rate is adjusted using colloidal solids and organic polymers to reduce loss of filtrate to the formation and prevent buildup of a thick filter cake which would restrict the wellbore. Excessive decrease of the filtration rate can be costly and may result in a viscous fluid that may affect the drilling rate adversely.

Large solid particles of various sizes and shapes may be added to control circulation loss where natural or induced fractures, highly porous formations, or vugular zones, ie, formations containing small cavities (vugs) larger than the matrix grain size, are encountered. Particle sizes from those large enough to bridge the opening down to those fine enough to seal small spaces between the larger bridging particles may be required to prevent drilling fluid losses into these zones.

Water Chemistry. Water is present in all but purely gaseous or oil drilling fluids, both of which comprise only a small percentage of drilling fluid applications. The water may be present as fine droplets in a mist, emulsified in an organic continuous phase, or as is most common, comprise the continuous phase of the drilling fluid. Water added to drilling fluids may be fresh water, seawater, or saturated salt solutions. The salt in the last is normally sodium chloride but may be another halide or alkali or alkaline-earth salt. The nature of the dissolved salts affects colloidal clays and other additives and thus must be monitored together with properties such as salinity, total hardness (calcium plus magnesium), pH, and alkalinity. Drilling fluids are nearly always basic. The pH ranges from 6 to 13 depending on the type of system. Concentrations of soluble carbonates, sulfide, sulfite, etc, may also be determined. These ions may be added intentionally or incorporated during drilling.

Drilling Fluid Materials

Density Control. The pressure exerted by the column of drilling fluid in the well balances formation pressures to prevent uncontrolled influx of formation fluids which may result in a blowout. The mud density must be controlled accurately by suitable weighting materials that do not adversely affect the other properties. Most important is the specific gravity of the weighting agent as well as its water insolubility and chemical inertness. The weighting material should be ground to the preferred particle-size distribution and be relatively nonabrasive. Various finely ground, solid weighting materials that have found application in drilling are listed in Table 1. As of 1995, little weighting material other than barite was used in the United States.

Table 1. Solid Materials Used to Increase the Density of Drilling Muds

Material	CAS Registry Number	Formula	Specific gravity	Hardness, Mohs'	Charac-teristics	Advantages	Limitations
barite	[13462-38-2]	BaSO ₄	4.5 ^a , 4.2 ^b	2.5–3.5	whitegray to reddish	API ^c standard in the industry, readily available, low cost	soft mineral, can produce fines and increase viscosity, may contain trace metals
hematite	[1317-60-8]	Fe ₂ O ₃	4.9–5.3	5.5–6.5	iron oxide, impuri-ties; black to red depend-ing on particle size	API ^c high density, micaeous structure, ease of process-ing and mixing, HCl sol, low attri-tion rate	abrasive, undesir-able color
magnetite	[1309-38-2]	Fe ₃ O ₄	5.0–5.2	5.5–6.5	iron ore, often Ti and Mg; black mineral	high density; HCl sol; scavenges H ₂ S	abrasive, magnetic suscepi-bility
ilmenite	[12168-52-4]	TiO ₂ ·FeO	4.5–5.1	5–6	titanic iron ore, black mineral	high density; low attri-tion rate	not commer-cially available; abrasive
siderite	[14476-16-5]	FeCO ₃	3.7–3.9	3.5–4	spathic iron ore, various colors	acid sol, mainly used in comple-tion and workover fluids	low density, not widely available
dolomite	[16389-88-1]	CaCO ₃ ·MgCO ₃	2.8–2.9	3.5–4	carbonate of calcium and mag-nesium	acid sol, used in comple-tion and workover fluids	low density
calcite	[13397-26-7]	CaCO ₃	2.6–2.8	3	limestone, occurs in sedimen-tary rock, oyster shells	readily acid sol, available in range of particle sizes, primarily used in comple-tion and workover fluids	low density
sodium chloride	[7647-14-5]	NaCl	2.165	2	cubic structure, evaporite	water soluble, available in range of particle sizes, primarily used as bridging solids	requires salt-satur-ated fluid, low density

^a Value is for pure material.
^b Value is for API grade.
^c API specifications available.

Barite, predominately BaSO₄, meets the overall requirements for weighting material better than other materials and is used for increasing the density of drilling fluids throughout the world. Commercial barite has a lower specific gravity than pure barium sulfate owing to the presence of associated minerals, such as silica. Barite is virtually insoluble in water and does not react with other mud constituents. Most operators prefer barite that meets API specifications (Table 2) (23). The barite content in mud depends on the desired density but can be as high as 2000 kg / km³ (700 lb / bbl).

Table 2. API Specifications for Barite and Hematite

Assay	Barite ^a	Hematite ^b
specific gravity ^c	4.20	5.05
wet-screen analysis, % residue ^d		

>75 μm	3.0	1.5
>45 μm		15
particles <6 μm , % ^d	30	15
soluble alkaline-earth metals as calcium, mg/kg ^d	250	100

^a Test equipment and procedures as per Ref. 23.

^b Test equipment and procedures as per Ref. 24.

^c Value given is minimum.

^d Values given are maximum.

In the early 1980s, an anticipated worldwide barite shortage led to development of alternative materials for weighting drilling fluids. The technology favors systems having the lowest possible solids content because of the resulting higher penetration rate, easier control of mud properties, and fewer problems experienced during drilling. Alternative weighting materials having specific gravities higher than that of barite offer this advantage. The only other high specific gravity material that is used to any degree as of 1995 is hematite. Ilmenite, used for a short time, is no longer available commercially (25).

The specifications for drilling fluid hematite have been set by the API and are listed in Table 2 (24). Hematite is used most frequently in high density oil-based muds to minimize the total volume percent solids (26). The abrasivity of hematite limits its utility in water-based muds.

Calcite and siderite (27) are used occasionally because of their solubility in hydrochloric acid which offers a method of removing mud filter cake deposited on productive formations. Calcite and siderite are used most frequently in workover or completion fluids when a nondamaging fluid is required, ie, one that can be removed by acidizing at a later time.

Solid salt, ground and packaged in several particle size grades, can be used in saturated salt brines to increase the fluid density (28). However, sized salt is most often used as a water-soluble material for bridging or sealing porous formations. At one time the sized salt systems were used primarily for completion or workover operations, but use has increased as drill-in fluids for horizontal wells (29).

Solids-free fluids are used occasionally to achieve high drilling rates and for workover and completion operations to avoid damage arising from particulate invasion of the productive formations. Weighted fluids without solids are provided by solutions of various water-soluble salts. Aside from the density required by a specific application, cost and corrosion have to be considered. Brine sources include seawater, natural brine, and manufactured salts. The maximum densities obtainable using soluble salts that have been used as drilling, completion, and workover fluids are listed in Table 3. The desired density may be achieved by mixing one or more salts. Densities to 2300 kg/m³ (19.2 lb/gal) have been obtained using mixtures of calcium chloride, calcium bromide, and zinc bromide. Slightly higher densities could be reached with near-saturated solutions of cesium formate (2367 kg/m³ (19.7 lb/gal)) (30). Such high density salt solutions are extremely expensive. Sodium chloride and calcium chloride are the only salts that are used in large quantities for drilling (see CALCIUM COMPOUNDS; SODIUM COMPOUNDS).

Table 3. Soluble Salts that Increase the Density of Drilling Mud and Workover and Completion Fluids

Salt	CAS Registry Number	Formula	Specific gravity	Saturated or near saturated brine, 20–25°C	
				Composition, wt %	Density, g/cm ³
potassium chloride	[7447-40-7]	KCl	1.984	26.5	1.18
sodium chloride	[7647-14-5]	NaCl	2.165	28.5	1.20
calcium chloride	[10043-52-4]	CaCl ₂	2.150	46	1.46
zinc chloride	[7646-85-7]	ZnCl ₂	2.910	81	2.14
sodium bromide	[7647-15-6]	NaBr	3.203	45	1.50
calcium bromide	[7789-41-5]	CaBr ₂	3.353	59	1.70
zinc bromide	[7699-45-8]	ZnBr ₂	4.201	82	2.65
potassium formate	[590-29-4]	KCHO ₂	3.48	76	1.60
sodium formate	[141-53-7]	NaCHO ₂	1.92	45	1.34
cesium formate	[3495-36-1]	CsCHO ₂ ·H ₂ O	3.372	83	2.37

The chemical and mechanical dispersion of the drilled solids tends to increase the percentage of small-sized solids in a mud as drilling progresses. The incorporation of a limited amount of drilled solids (several volume percent) is an economical way of increasing the density of low density muds, but it also reduces penetration rates; hence, drilled solids are usually kept to a minimum. The common clay and formation solids encountered in normal drilling operations are as follows:

Material	Specific gravity
dolomite	2.8–3.2
limestone	2.8–3.0
feldspars	2.6–2.7
sandstone	2.3–2.6
clay	2.3–2.6
bentonite	2.3–2.4
salt (NaCl)	2.16
cement	1.6–2.0
coal	1.35

The cement is encountered when drilling a cement plug or out of new casing.

Viscosity Buildup. The drilling fluid removes cuttings from the wellbore as drilling progresses. This process is governed by the angle of the hole and the velocity at which fluid travels up the annulus, as well as by the fluid viscosity or flow properties, and fluid density. The cuttings removal efficiency usually increases with increasing viscosity and density, although at high wellbore angles a less viscous fluid may be desirable provided high flow rates can be achieved. Viscosity depends on the concentration, quality, and state of dispersion of suspended colloidal solids.

Although numerous mud additives aid in obtaining the desired drilling fluid properties, water-based muds have three basic components: water, reactive solids, and inert solids. The water forming the continuous phase may be fresh water, seawater, or salt water. The reactive solids are composed of commercial clays, incorporated hydratable clays and shales from drilled formations, and polymeric materials, which may be suspended or dissolved in the water phase. Solids, such as barite and hematite, are chemically inactive in most mud systems. Oil and synthetic muds contain, in addition, an organic liquid as the continuous phase plus water as the discontinuous phase.

The most important commercial clays used for increasing the viscosity of drilling fluids are bentonite [1302-78-9], attapulgite [1337-76-4], and sepiolite [15501-74-3]. For oil-base and synthetic-base muds, organophilic clays are used. These clays produce viscosity and help suspend weighting materials. The active ingredient in bentonite is a smectite, montmorillonite [1318-93-0], having a three-layer plate-shaped crystalline structure. The three-layer sheets or platelets consist of a middle octahedral alumina layer and two outer tetrahedral silica layers. Because of lattice defects in the alumina, and less often in the silica layers, the flat planar surfaces are negatively charged and have associated cations (primarily sodium and calcium) to achieve electroneutrality. Bentonite hydrates in the presence of fresh water and disperses to varying degrees, depending on the nature of the cations that are loosely held and exchangeable. Hydration and dispersion are enhanced by the presence of sodium ions in the exchange or planar surface positions. Divalent cations such as calcium reduce the degree to which the individual platelets disperse (see CLAYS).

As hydration of bentonite proceeds in fresh water, the individual platelets separate and eventually form a stable colloidal dispersion that is stabilized by electrical interactions between clay platelets (31). This colloidal system has highly non-Newtonian rheological properties and the suspending and shear-thinning characteristics desired in a drilling fluid. The hydration and dispersion of bentonite clay, and thus the viscosity of the final system, can be altered considerably by the presence of electrolytes in the water, exchange reactions that convert the clay from a high swelling form (sodium) to a low swelling form (such as calcium), or adsorption of polyelectrolytes which prevent separation of the clay platelets.

Bentonite concentrations in drilling fluids vary widely, but may range up to 100 kg m³ (35 lb bbl). Specifications for the three grades of drilling fluid bentonite recognized by the API are listed in Table 4 (32). The higher performance grades are produced mainly in the Wyoming–Montana–South Dakota area. This bentonite contains montmorillonite clay in both sodium and calcium forms. The sodium form predominates. Most such bentonites are processed using a small amount of a peptizing polymer to enhance the viscosity building properties of the clay. This is the standard API bentonite. High quality bentonite that is not treated in any way to enhance its viscosity building characteristics is sold as API nontreated bentonite.

Table 4. API Specifications for Bentonite Clay^a

Assay	Bentonite	Nontreated bentonite	OCMA bentonite
bentonite content in 350 cm ³ distilled water, g	22.5	25.0	22.5
suspension properties			
viscometer dial reading at 600 rpm ^b	30	10 ^c	30
yield point plastic viscosity ratio ^d	3	1.5	6
filtrate volume ^d , cm ³	15.0	12.5	16.0
residue, % >75 μm ^d	4.0		2.5
moisture, % ^d	10.0		13.0

^a Test equipment and procedures as per Ref. 32.
^b Value given is minimum.
^c Value is dispersed plastic viscosity in mPa·s(cP).
^d Value given is maximum.

Large deposits of lower yielding calcium bentonites are found in the coastal plains of the Gulf of Mexico and in many areas around the world. This material can be used in drilling muds if upgraded with sodium carbonate and peptizing polymers such as polyacrylates (33). The name API OCMA bentonite identifies such bentonites for drilling fluid use. OCMA is a term left over from the defunct organization, the Oil Companies Materials Association, that until the early 1980s set specifications for drilling fluid materials used outside the United States. The API, which has adopted upgraded versions of many of the old OCMA specifications, is the specification-setting organization as of the mid-1990s.

Fast-yielding bentonite treated with high molecular weight polyacrylamide and polyacrylate polymers is also available commercially. It produces roughly the same viscosity as twice the amount of untreated bentonite. Additions of small amounts of certain polyacrylamides, called bentonite extenders, to a freshwater mud containing bentonite give the same effect. Drilling fluids formulated with bentonite and an extending polymer in this manner are called low solids, nondispersed muds (34).

Another bentonite extender, which also enhances the viscosity of attapulgite slurries, is a mixed metal (magnesium and aluminum) layered hydroxide [7732-18-5], usually called MMH (35). This material is most frequently used to increase the viscosity sufficiently that metal cuttings can be removed from the well during milling operations. MMH-treated muds are also finding application in horizontal wells. A mixed metal (calcium, magnesium, and aluminum) silicate [1327-39-5] (MMS), is being applied in a similar way.

When the water phase of the drilling fluid contains substantial amounts of electrolyte, salt water clays such as attapulgite and sepiolite are added to raise viscosity. Attapulgite is used solely for its suspending qualities. It has a fibrous texture and crystalline, needle-like, hydrated magnesium silicate particles. The crystal structure is that of a double chain of silicon and oxygen linked by magnesium and calcium (36). Attapulgite clays increase viscosity regardless of the composition of the makeup water. This ability does not depend on hydration, but rather on the extent to which the bundles of needles are broken up by a shearing force.

Sepiolite is a hydrated magnesium silicate that contains less substituted aluminum than attapulgite, which it closely resembles. Sepiolite occurs in fibrous and elongated lath-like particles. It is stable at higher temperatures than attapulgite (37,38) and therefore is used in geothermal drilling fluids (see GEOTHERMAL ENERGY). The properties of attapulgite and sepiolite specified by the API for drilling fluid use are as follows (39).

Assay	Value
clay concentration in 350 cm ³ saturated salt water, g	20
minimum viscometer dial reading at 600 rpm	30
maximum residue, % >75 μm	8.0
maximum moisture, %	16.0

Oil-dispersible or organophilic clay provides viscosity and suspending qualities in oil-based muds. It is prepared from bentonite, hectorite [12173-47-6], or attapulgite and aliphatic amine salts. The products obtained from amines having 12 or more straight-chain carbon atoms swell and form gels in hydrocarbon fluids (40–42). The amino groups replace the sodium and calcium originally present on the clay surface. Oil-dispersible clays can suspend solids in oil without requiring additional soaps and emulsifying agents. Addition levels of organophilic clays are ca 3–11 kg m³ (1–4 lb bbl) in diesel oil muds and ca 8–23 kg m³ (3–8 lb bbl) in mineral oil and synthetic-base muds.

A wide variety of organic polymers serve a number of useful purposes in drilling fluids, the most important of which are to increase viscosity and control filtration rates (Table 5). These polymers are either natural polysaccharides, eg, starch [9005-25-8] (qv), guar gum, xanthan gum [11138-66-2], and other biopolymers (see GUMS; MICROBIAL POLYSACCHARIDES); or derivatives of natural polymers, eg, cellulose (qv), lignosulfonate, and lignite and synthetic polymers, eg, polymers and copolymers of acrylic acid, acrylonitrile, acrylamide, and 2-acrylamido-2-methylpropanesulfonic acid (AMPS). The most commonly used polymeric viscosity builders are the celluloses, xanthan gum, and polyacrylamides.

Table 5. Polymers Used in Water-Base Drilling Fluids

Polymer	Chemical description	Ionic state	Thermal degrad-ation, °C	Application	Limitations
Fermentable					
starch	polysaccharide: amylose, amylopectin	nonionic	110	filtration control	fermentable at low pH and salinity
xanthan gum	microbial polysaccharide	anionic	140	viscosity builder	poor yield in calcium brines
hydroxy-propyl guar	polysaccharide: galactose, mannose	nonionic	95	viscosity builder	low tem-perature stability
Unfermentable					
CMC PAC	sodium carboxy-methyl cellulose	anionic	140	filtration control, viscosity builder	sensitive to salinity, multivalent ions
HEC	hydroxyethyl cellulose	nonionic	110	viscosity builder, acid degradable	primarily for completion workover fluids
polyacrylates and poly-acryl-amides	sodium polyacrylate, polyacryl-amide, polyacrylic acid-co-acrylamide	mixed	200	dispersant, filtration control, viscosity builder, shale stabilizer	poor perfor-mance in presence of multivalent ions
lignosulfonate	sulfonated lignin, cations vary ^a	anionic	175	deflocculant, filtration control	may contain heavy metals, requires alkaline conditions
tannin	quebracho, may be sulfomethyl-ated, may contain a heavy metal	anionic	200	deflocculant	more effective if sulfo-methylated, wide pH range
lignite	leonardite	anionic	205 +	filtration control, deflocculant	requires alkaline conditions, sensitive to salinity and multivalent ions

^a Cations are often chromium.

Sodium carboxymethyl cellulose [9004-32-4] (CMC) and hydroxyethyl cellulose [9004-62-0] (HEC) are the cellulosics most widely used in drilling fluids (43). CMC is manufactured by carboxymethylation of cellulose which changes the water-insoluble cellulose into the water-soluble CMC (44). Hydroxyethyl cellulose and carboxymethyl hydroxyethyl cellulose (CMHEC) are made by a similar process. The viscosity grade of the material is determined by the degree of substitution and the molecular weight of the finished product.

The effectiveness of sodium carboxymethyl cellulose, an anionic polymer, as a viscosity builder decreases with increasing electrolyte concentration. This polymer can be coprecipitated with calcium and magnesium by raising the pH of the mud. The polyanionic cellulose (PAC), which has a higher degree of substitution than CMC, was introduced to overcome some of these limitations (43). CMC and PAC are available in several viscosity and purity grades. Concentrations are ca 0.6–14 kg m⁻³ (0.2–5 lb bbl). API specifications for high viscosity, technical-grade CMC for use as a viscosity builder are listed in Table 6 (45). The primary application of both CMC and PAC is in the control of filtration rates.

Table 6. API Specifications for Technical-Grade High Viscosity CMC^a

Solvent system	CMC concentration, g 350 cm ³	Minimum viscometer dial reading at 600 rpm	Maximum filtrate volume, cm ³
distilled water	2.20	30	
40 g L salt water	2.70	30	
saturated salt water	2.50	30	
slurry ^b	3.15		10

^a Test equipment and procedures as per Ref. 45.

^b Slurry contains 35 g of API Standard Evaluation Base Clay per 350 cm³ saturated salt solution.

Hydroxyethyl cellulose (HEC), a nonionic thickening agent, is prepared from alkali cellulose and ethylene oxide in the presence of isopropyl alcohol (46). HEC is used in drilling muds, but more commonly in completion fluids where its acid-degradable nature is advantageous. Magnesium oxide stabilizes the viscosity-building action of HEC in salt brines up to 135°C (47). HEC concentrations are ca 0.6–6 kg m⁻³ (0.2–2 lb bbl).

Xanthan gum is a high molecular weight microbial polysaccharide produced by the bacterium *Xanthomonas campestris* (48). Commercially xanthan gum is produced by a fermentation (qv) process and precipitation of the gum in alcohol. It is a viscosity builder and suspending agent and can be used in almost any type of water (49). Although xanthan gum solutions support bacterial growth, a preservative is usually not needed unless the solution is to be stored. Because of suspending ability at low concentrations and in electrolyte solutions, xanthan gum is widely used for drilling, workover, and completion fluids. Concentrations range from ca 0.6 to 6 kg m⁻³ (0.2–2 lb bbl). Two other biopolymers, succinoglucan gum [39464-87-4] (50) and welan gum [96949-22-3] (48,51), are also finding some use in drilling fluids at concentrations similar to xanthan gum.

Guar gum is a nonionic, branched-chain polysaccharide, a galactomannan that is usually hydroxypropylated for use in drilling (52). It produces viscous solutions in fresh or salt water at concentrations of ca 3–6 kg m⁻³ (1–2 lb bbl). It is used in solids-free and low solids muds and degrades rapidly above 80°C, limiting its use to shallow wells.

High molecular weight polyacrylamides are used as viscosity builders in freshwater muds (53) or as bentonite extenders. The ionic nature of the polyacrylamide may range from nonionic to anionic (30% hydrolyzed) depending on the situation. Molecular weights ranging from >3 × 10⁶ are used for this purpose. Polymer concentrations of 0.7–2.8 kg m⁻³ (0.25–1.0 lb bbl) are used depending on the application.

Occasionally polymers are used to increase the viscosity of oil-base and synthetic-base muds. The polymers for this use are typically sulfonated polystyrenes or ethylene–propylene terpolymers (EPDM) (54,55). Such polymers are usually used in conjunction with an organophilic clay.

Viscosity Reduction. Proper control of viscosity and gel strengths is essential for efficient cleaning of the borehole, suspension of weight material and cuttings when circulation is interrupted, and to minimize circulating pressure losses and swab surge pressures owing to axial movement of the drill string. Viscosity may be increased as previously indicated, but there is often the necessity of reducing the viscosity. A reduced viscosity can be achieved

by thinning or deflocculating clay–water suspensions. Thinning is measured as a reduction of plastic viscosity, yield point, or gel strength, or a combination of these properties. Viscosity is reduced by decreasing the solids content and the number of particles per unit volume, or by neutralizing the attractive forces between particles. Although mud thinners or viscosity-reducing chemicals are added to reduce flow resistance and gel development, reduction of filtration rates and filter cake thickness and stabilization of mud properties at elevated temperatures often results. Typical mud-thinning chemicals are polyanionic materials that are adsorbed on positive edge sites of the clay particles, thereby reducing the attractive forces between the particles without affecting clay hydration (31,56).

Thinners or deflocculants for clay–water muds include polyphosphates, tannins, lignites, lignosulfonates, and low molecular weight polyacrylates and their derivatives. These materials also can remove chemical contaminants by precipitation or chelation. Sodium polyphosphates are effective deflocculants for clays in fresh water and were among the first thinners used in drilling fluids (57). These are not effective in muds having salinities >10 g L. Polyphosphates effectively soften hard water by forming insoluble complexes with calcium and magnesium ions. The reversion of the polyphosphates to orthophosphates occurs rapidly as the temperature approaches 95°C and may cause thickening of the mud. Table 7 gives the four most commonly used phosphates. The application of polyphosphates is limited to temperatures below 80°C, and supplemental control along with organic thinners. Concentrations of polyphosphates of ca 0.3–3 kg m³ (0.1–1 lb bbl) are typical.

Table 7. Phosphate Drilling Mud Thinners

Compound	CAS Registry Number	Molecular formula	pH of aqueous soln	Observations
sodium acid pyrophosphate (SAPP)	[7758-16-9]	Na ₂ H ₂ P ₂ O ₇	ca 4.2	for cement contamination
sodium tetraphosphate	[7727-67-5]	Na ₄ P ₄ O ₁₃	ca 7.5	generally preferred
tetrasodium pyrophosphate (TSPP)	[7722-88-5]	Na ₄ P ₂ O ₇	ca 10	
sodium hexametaphosphate	[10124-56-8]	(NaPO ₃) ₆	ca 7	glassy material

Tannins occur in many plants and are separated by extraction. At present, only quebracho extract is used as a mud thinner in significant quantity in the United States. Quebracho is an acidic material and performs best at high pH. It is an excellent thinner for lime-treated and cement-contaminated muds. However, it is not effective at high salt concentrations. Sulfomethylated tannin products are functional over a wide range of pH and salinity and have either been treated with chromium for good thermal stability (58) or are chrome free. Concentrations of tannin additives are ca 1.5–18 kg m³ (0.5–6 lb bbl).

Lignite, generally leonardite, and lignite derivatives are applied in water-based muds as thinners and filtration control agents. Leonardite is an oxidized lignite having a high content of humic acids, which may be described as carboxylated phenolic polymers (59,60). Little is known about the chemical structure.

Natural lignite is not a good thinner for low pH water-based muds at moderate temperatures, but can be an excellent thinner or mud conditioner at high pH and high temperatures where the humic acids are solubilized. It has better temperature stability than most plant tannins or lignins. The performance of lignite can be greatly improved by solubilizing processes such as caustic treatment, sulfonation, sulfomethylation, or chelation with heavy metals. Lignite concentrations in drilling fluids are ca 3–60 kg m³ (1–20 lb bbl).

Lignosulfonate thinners are among the most versatile and important chemicals used in water-based drilling fluids. There have been many studies on the chemical and physical characteristics of lignosulfonates, but their structure has not been clearly defined (61). Lignosulfonates are by-products of the sulfite process for separating cellulose pulp from wood (qv). These can be described as polyanionic natural polymers. Their mud-thinning quality depends on the source of the lignin, degree of sulfonation, molecular weight distribution, impurities, and the amount and type of chelated heavy metals (62). Commercial chrome lignosulfonate [9066-50-6] contains up to ca 3.5–4.5 wt % trivalent chromium. Ferrochrome lignosulfonate [8075-74-9] contains an additional 1–2 wt % iron. A number of chrome-free lignosulfonates are also available. Cations associated with the lignosulfonate molecule may include sodium, potassium, calcium, titanium–zirconium (63), among others. The chrome-free lignosulfonates may not be as effective overall as those that contain some trivalent chromium, but are useful where chromium is not desirable for environmental reasons (see LIGNIN).

Chrome and ferrochrome lignosulfonates are effective deflocculants in most water-based muds over a wide range of salinity, hardness (qv), and pH (8.5–12.5). Lignosulfonates at high concentrations provide some filtration control and inhibit disintegration and dispersion of shale cuttings. Thinning effectiveness declines at 120–150°C, although field experience confirms that muds containing chrome lignosulfonate can be used at bottomhole temperatures of >175°C. The effectiveness of lignosulfonates at high temperatures depends on the solids content and the type and concentration of electrolytes. In varied uses, the lignosulfonates have concentrations of 3–43 kg m³ (1–15 lb bbl).

Low molecular weight (1000–5000) polyacrylates and copolymers of acrylic acid and AMPS are used as dispersants for weighted water-base muds (64). These materials, 40–50% of which is the active polymer, are usually provided in a liquid form. They are particularly useful where high temperatures are encountered or in muds, which derive most of their viscosity from fine drill solids, and polymers such as xanthan gum and polyacrylamide. Another high temperature polymer, a sulfonated styrene maleic–anhydride copolymer, is provided in powdered form (65,66). All of these materials are used in relatively low (ca 0.2–0.7 kg m³ (0.5–2 lb bbl)) concentrations in the mud.

A hydrolyzed cereal solid, predominately a hexasaccharide, is used in high pH lime muds for reducing the yield point and gel strength (67). This additive has been used in systems treated with both sodium hydroxide and potassium hydroxide in addition to other additives common to lime muds (68). A second viscosity-reducing additive used in lime muds is a graft copolymer of acrylic acid and calcium lignosulfonate (69). Both of these materials are used at levels of 6–17 kg m³ (2–6 lb bbl).

Filtration Control. Filtration control is particularly important in permeable formations where the mud hydrostatic pressure exceeds the formation pressure. Proper filtration control reduces drill-string sticking and drag, and rotary torque, as well as minimizing damage to protective formations; in some formations it improves borehole stability. Several types of materials are available for water-based muds and application varies according to the type and the chemical environment of the mud. These include clays, organic polymers, and lignite derivatives. The bentonite present in the system often acts as the primary filtration control agent. It not only develops viscosity, but also lowers the filtration rate, particularly in freshwater muds. The ability of bentonite clay to control filtration is attributed to the flat, plate-like particle shape, the capacity to disperse and hydrate, the ability to form a compressible filter cake, and the colloidal to near-colloidal particle size.

Hydrating bentonite in fresh water before adding it to the mud greatly increases its efficiency when the makeup water is contaminated with salt and or hardness. Prehydrated bentonite can be protected from dehydration by lignosulfonate (70) or sulfomethylated tannin when used in saturated salt water. Salt water clays, such as sepiolite and attapulgite, provide no filtration control and are normally used with suitable filtration control agents.

Although a combination of bentonite clay and an organic thinner provides filtration control in many water-based muds, additional control generally is needed. Filtration additives for both fresh- and salt water muds are usually organic polymers and lignites (see Table 5). Cost and viscosity-building characteristics should be considered, particularly in weighted high solids muds.

Starches, used first in the late 1930s for filtration control (71), are still in use in the 1990s. Corn starch is most commonly used in the United States. Potato starch is also used in the United States but primarily in Europe and elsewhere. Both corn and potato starches are pregelatinized before dispersion in water (72). The API specifications for drilling fluid starch are listed in Table 8 (73).

Table 8. API Specifications for Starch^a

--	--	--	--	--	--

Starch slurry ^b solvent	Viscometer dial reading at 600 rpm ^c	Maximum filtrate volume, cm ³
40 g L salt water	18	10
saturated salt water	20	10

^a Test equipment and procedures as per Ref. 73.
^b Slurry contains 3.5 g starch and 35 g of API Standard Evaluation Base Clay per 350 cm³.
^c Values given are maximum.

Starch is subject to fermentation by many microorganisms and, unless the mud is saturated with salt or the pH is >11.5, a preservative or biocide must be added if the mud is to be used for an extended period of time. The most common biocide until the mid- 1980s was paraformaldehyde [9002-81-7]. This material has been largely replaced by isothiazolones (at 5–10 ppm conc) (74), carbamates, and glutaraldehyde [111-30-8]. Alternatively, the biocide may be incorporated during the processing of the starch and is present in the commercial product.

Starch is degraded by heat and agitation. Under continued circulation at >95°C, starch can break down rapidly. Its performance also is affected by pH changes in the presence of calcium and magnesium ions. Starch coprecipitates with calcium when caustic soda is added to mud containing dissolved calcium salts. Starch molecules are hydrolyzed at very low or high pH in the presence of calcium and magnesium above 110°C. An acidic environment also may prevent hydration and thus limit the effectiveness of starch as a filtration control agent.

Numerous modifications and derivatives of starch have been made for application in drilling and workover fluids (75). Most modified starches are cross-linked to some degree to improve thermal stability. Carboxymethyl and hydroxypropyl starches are finding increasing application in drilling fluids. A nonfermenting product has been prepared by blending moist starch with 3 wt % bis(2-hydroxy-3,5-dichlorophenyl) sulfide and passing the mixture under pressure through a heated extruder (76–78). This product causes a smaller increase in drilling fluid viscosity than the usual pregelatinized starch and is reported to be effective in retarding shale disintegration and dispersion (79). There are some applications involving cationic starches as well, although toxicity must be considered for offshore use (76). The concentration of starch and its modified forms in muds ranges from 6 to 29 kg m³ (2–10 lb bbl).

Carboxymethyl cellulose (CMC) and polyanionic cellulose (PAC) are available in several viscosity grades and high and low purity grades. All grades can be effective filtration control agents depending on the well conditions. The effectiveness of CMC in reducing filtration decreases with increasing salt concentration. A polyanionic cellulose (PAC) polymer is available that is designed particularly for application in muds of high salt concentration (43,80). Most CMC and PAC materials are sodium salts but there are also potassium versions to take advantage of the shale stabilizing properties of this ion (81). The API specifications for low viscosity technical-grade CMC used for filtration control (82) state that for a 30 g L solution, the maximum viscometer dial reading at 600 rpm should be 90 and for a 9.0 g L solution of CMC in a slurry containing 35 g of API Standard Evaluation Base Clay per 350 cm³ saturated salt solution, the maximum filtrate volume should be 10 cm³.

Although not designed to control filtration, HEC may be effective as a filtration control agent in combination with other organic polymers in waters having salinities up to saturation.

Acrylate and acrylamide polymers have several uses in drilling fluids, one of which is for filtration control. Sodium polyacrylates [9003-04-7] having molecular weights near 250,000 are excellent temperature-stable filtration control agents for both fresh- and salt water muds, provided the concentration of water-soluble calcium is <400 mg L (83). The calcium ions are precipitated using a carbonate such as soda ash, before adding the polyacrylate at concentrations up to ca 6 kg m³ (3 lb bbl).

Lignite products, mined, ground, and possibly treated with sodium or potassium hydroxide, are economical filtration control additives for some water-based muds, in addition to improving flow properties (84). A sulfonated lignite complexed with a sulfonated phenolic resin is an effective high temperature, filtration control additive, for both fresh- and seawater muds that contain high concentrations of soluble calcium, and does not affect viscosity (85,86). A high molecular weight polyanionic lignin has also found application for high temperature muds with a high electrolyte content (87). These and similar products can provide filtration control for high density, high solids muds above 120°C.

A number of synthetic polymers having the ability to control filtration rates at high temperature and in the presence of calcium and magnesium have also been developed (88). Such materials include vinyl sulfonate–vinyl amide copolymers (89,90), a copolymer of AMPS and N,N-dialkyl (meth) acrylamide (91) and a sulfonated hydroxylated polymer (92). Application levels for these materials range from 5 to 18 kg m³ (2–6 lb bbl). Sulfonated asphalt is also used for high temperature filtration control.

Filtration control in oil- and synthetic-base fluids is achieved by the emulsified aqueous phase, by the emulsifier package, and by additions of powdered solid materials. The powdered solids used for this purpose consist of asphalt [8052-42-4], gilsonite [12002-43-6], and amine-treated lignite. Addition levels range from 6 to 29 kg m³ (2–10 lb bbl). The choice of additive depends on the nature of the base fluid, the emulsifier package, and most importantly the downhole temperature. Asphalt and gilsonite of various softening points can be selected based on temperature. Styrene–butadiene copolymers in the form of aqueous dispersions have also found application as filtration control agents in oil-base muds.

Cellulosic fibers, powdered limestone, gilsonite, and asphalt are frequently added to both water and oil muds at levels of 10 to 25 kg m³ (4–10 lb bbl) when high differential pressures are encountered to control seepage losses to the formation. This treatment also is used to improve the quality of the mud filter cake to reduce the chance of differential pressure sticking.

Alkalinity Control. Water-base drilling fluids are generally maintained at an alkaline pH. Most mud additives require a basic environment to function properly and corrosion is reduced at elevated pH. The primary additive for pH control is sodium hydroxide [1310-73-2] in concentrations from 3 to 14 kg m³ (1–5 lb bbl).

The second most common alkalinity control agent is lime [1305-78-8], normally in the form of calcium hydroxide [1305-62-0], used in both water and oil muds. In the latter, the lime reacts with added emulsifiers and fatty acids to stabilize water-in-oil emulsions. Lime is used in brine systems containing substantial quantities of soluble calcium and in high pH lime muds. Concentrations are ca 6–57 kg m³ (2–20 lb bbl) (see LIME AND LIMESTONE).

Potassium hydroxide [1310-58-3] is occasionally used for alkalinity control. This is particularly true for some polymer and lime muds where a low sodium level is desired. The potassium level of such muds is quite low but has been attributed by some to provide stability to water-sensitive shale formations (68,93).

A fourth alkalinity control additive is magnesium oxide [1309-48-4], which is used in clay-free polymer-base fluids (47). Magnesium oxide provides an alkaline environment and, as it is only slightly soluble, also has a buffering effect. It enhances the thermal stability of polymer solutions by preventing a pH decrease to neutral or slightly acidic conditions at elevated temperatures. It is mainly applied in completion or workover operations where clay-free acid-soluble fluids are desired.

There are occasions where the mud pH must be lowered such as after drilling fresh cement or overtreatment by one of the alkaline materials discussed. Organic acids that have been used for this purpose include acetic acid [64-19-7], citric acid [77-92-9], and oxalic acid [144-62-7]. These materials are used infrequently. Inorganic additives used to lower pH levels include sodium bicarbonate [144-55-8] and sodium acid pyrophosphate [7758-16-9] (SAPP). Of the two, sodium bicarbonate is used the most by far.

Removal of Contaminants. A drilling fluid contaminant is any material or condition encountered during drilling operations that adversely affects the performance of the fluid. Elevated temperatures and drill solids are encountered in every drilling operation. In most wells these are handled easily, but in some wells one or both can seriously reduce drilling efficiency. Temperature problems normally are treated using viscosity or filtration control additives, material having better thermal stability, or possibly by replacement of the mud system with an oil or synthetic mud. Drill solids are removed mechanically by various combinations of screens, hydrocyclones, and centrifuges, or chemically by flocculants. Dilution or replacement of part or all of the mud system may reduce drill solids to tolerable levels.

Various inorganic chemicals remove soluble contaminants encountered during drilling. Salt, NaCl, is a common contaminant that can be removed only by dilution. The adverse effects of salt, primarily clay flocculation, can be overcome by a deflocculant such as a lignosulfonate or sulfomethylated

tannin. Soluble calcium from cement, gypsum, or anhydrite is frequently encountered and has a flocculating effect on drilling clays. Calcium is removed using a phosphate, sodium carbonate [497-19-8], sodium bicarbonate (particularly in the case of cement contamination), and occasionally oxalic acid.

If calcium problems persist, the mud may be converted to a calcium-based system by adding a calcium source, such as gypsum [13397-24-5] or lime. Gypsum and lime are also used to control bicarbonate carbonate ions that can cause mud gelation and rheological problems, particularly in high density muds in deep, hot wells (94). Carbonate sources include carbon dioxide-bearing zones and soluble carbonate formations. Carbon dioxide is also generated by some mud additives and by excessive treatment for removal of calcium.

Stabilization of Water-Sensitive Formations. Many subsurface formations encountered during drilling are water-sensitive shales containing various amounts of clay minerals. The clay mineral components may include a highly swelling smectite or less water-sensitive illite, mixed layer smectite–illite, kaolinite, or chlorite. All shales appear to swell to some extent when contacted by fresh water. Those containing smectite or mixed layer clays are much more sensitive than illitic shales. The uptake of water by shales has two effects: a volume change owing to swelling, and a strength reduction as the water content increases. This may result in flow of plastic shale into the wellbore, softening and erosion of the exposed shale, or spalling of hard shale, all of which can cause expensive operational problems. In the latter case, large hard pieces of formation fall around the drill string, the removal of which may become difficult.

A variety of methods have been devised to stabilize shales. The most successful method uses an oil or synthetic mud that avoids direct contact between the shale and the emulsified water. However, preventing direct contact does not prevent water uptake by the shale, because the organic phase forms a semipermeable membrane on the surface of the wellbore between the emulsified water in the mud and the water in the shale. Depending on the activity of the water, it can be drawn into the shale (activity lower in the shale) or into the mud (activity higher in the shale) (95–97). This osmotic effect is favorable when water is drawn out of the shale; thus the aqueous phase of the oil or synthetic mud is maintained at a low water activity by adding a salt, either sodium chloride or more commonly, calcium chloride. The salt concentration is carried somewhat above the concentration required to balance the water activity in the shale to ensure water movement into the mud.

High initial cost and environmental restrictions prevent use of oil and synthetic muds in many cases where shale problems are expected. It is necessary then to treat a water-base mud to minimize the destabilizing effect of the drilling fluid. Salts, polymers, and other organic materials are added to the mud to reduce the water sensitivity of the shale, shale swelling, and weakening arising from mud contact, or the rate of water uptake by the shale.

Addition of a salt can transform the shale by cation exchange to a less sensitive form of clay, or reduce the osmotic swelling effect by reducing the water activity in the mud below that which occurs in the shale. These effects depend on the salt concentration and the nature of the cation. Salts containing sodium, potassium, calcium, magnesium, and ammonium ions are used to varying degrees.

Sodium chloride has long been used as a shale stabilizer because of low cost, wide availability, and its presence in many subsurface formations. The inhibitive nature of salt muds increases as the salt content increases from seawater to saturated sodium chloride. In addition to the sodium chloride consumed annually for drilling fluid, considerable quantities are incorporated while drilling salt zones. This material has been used more for minimizing washouts in salt zones than for stabilizing shales. High salt levels have found application in deep water drilling (7).

Calcium sources, such as gypsum and lime, promote cation exchange from sodium clay to a less-swelling calcium clay. Calcium concentrations are normally low (<1000 mg L) and osmotic swelling is only reduced if other salts are present. Calcium chloride has been used infrequently for this purpose but systems are available that allow high calcium chloride levels to be carried in the mud system (98).

A variety of shale-protective muds are available which contain high levels of potassium ions (10). The reaction of potassium ions with clay, well known to soil scientists, results in potassium fixation and formation of a less water-sensitive clay. Potassium chloride, potassium hydroxide, potassium carbonate [584-08-7] (99), tetrapotassium pyrophosphate [7320-34-5] (100), and possibly the potassium salts of organic acids, such as potassium acetate [127-08-2] (101) and formate, have all been used as the potassium source. Potassium chloride is generally preferred because of its low cost and availability. It is generally used in combination with a polymer or other organic material as discussed below.

Ammonium chloride [12125-02-9], ammonium sulfate [7783-20-2], and diammonium phosphate [7708-28-0] have also been used for shale stabilization (102,103). Ammonium ions have essentially the same effect on shales as potassium ions but use of ammonium salts is often objectionable because of the alkaline nature of the mud. In the North Sea and northern Europe, where magnesium-bearing salt formations are encountered, magnesium chloride [7786-30-3] is used, but in the United States it is used only on a small scale.

A number of nonionic and anionic polymers are employed in water-based muds to stabilize shales. These may be added to a freshwater mud or to a system containing one of the salts mentioned. Historically, high (>30 kg m⁻³) concentrations of chrome lignosulfonates were thought to stabilize hydratable shales. This technique rarely is used in the 1990s. Typically, shale-stabilization polymers include modified starches (78); cellulosic polymers such as CMC and HEC; gums such as guar, xanthan (104), and flax meal (105); and high molecular weight polyacrylamides of varying degrees of hydrolysis (9,53,106). A fluid containing a combination of potassium silicate [1312-76-1] and poly(vinyl alcohol) [9002-89-5] has also been used (107). The consumption of these materials for shale stabilization is difficult to estimate because they may also be used for viscosity or filtration control. Nonionic polymers are typically used at concentrations of 3 to 11 kg m⁻³ (1–4 lb bbl); anionic polymers can often be effective at much lower (0.5 to 3 kg m⁻³ (0.125–1 lb bbl)) concentrations.

The method of action of the polymers is thought to be encapsulation of drill cuttings and exposed shales on the borehole wall by the nonionic materials, and selective adsorption of anionic polymers on positively charged sites of exposed clays which limits the extent of possible swelling. The latter method appears to be true particularly for certain anionic polymers because of the low concentrations that can be used to achieve shale protection (8).

A number of cationic muds have been developed and used. These are formulated around quaternary amines or positively charged polymers (108,109). The polymer in some instances may be a cationic polyacrylamide. Poly(dimethylamine-*m*-epichlorohydrin) is another material that has been used successfully for drilling shale formations (110,111). Some of these additives may require a salt such as sodium or potassium chloride for best results.

A number of glycol and glycerol-base additives are being used to formulate shale protective muds usually in conjunction with a salt and or a polymer (112–115). The glycols (qv) consist of a number of combinations of ethylene oxide (qv) and propylene oxide depending on the purpose of the additive, not all of which are strictly for shale stabilization, and the desired cloud point. Two polyglycerol materials have been touted as having shale-protective properties (116–118). A low molecular weight poly(amino acid) has also been touted for its shale-stabilizing properties (119,120). An aluminum complex has also been used in combination with polymer muds for drilling shale (121).

A more recent addition to the list of shale protective water-base muds is a system developed around concentrated solutions of methyl glucoside [3149-68-6]. At concentration of 25% by weight and above, methyl glucoside appears to stabilize water-sensitive shales on par with a typical oil- or synthetic-base mud (122). Early field trials have been encouraging but much remains to be done before this material is considered a success (123).

Solid materials, such as gilsonite and asphalt, and partially soluble sulfonated asphalt may also be added to plug small fractures in exposed shale surfaces and thereby limit water entry into the formation (105,124). The asphalts are oxidized or treated to impart partial solubility. These materials may be softened by the downhole temperature, causing them to deform and squeeze into small openings exposed to the borehole. Laboratory tests designed to evaluate shale-stabilizing muds have confirmed the beneficial action of these materials (125) (see also SOIL STABILIZATION).

Surfactants. Surfactants (qv) perform a variety of functions in a drilling fluid. Depending on the type of fluid, a surfactant may be added to emulsify oil in water (o/w) or water in a nonaqueous liquid (w/o), to water-wet mud solids or to maintain the solids in a nonwater-wet state, to defoam muds, or to act as a foaming agent.

Lignites and lignosulfonates can act as o/w emulsifiers, but generally are added for other purposes. Various anionic surfactants, including alkylarylsulfonates and alkylaryl sulfates and poly(ethylene oxide) derivatives of fatty acids, esters, and others, are used. Very little oil is added to water-base muds in use offshore for environmental reasons. A nonionic poly(ethylene oxide) derivative of nonylphenol [9016-45-9] is used in calcium-treated muds (126).

Emulsifiers are incorporated in oil and synthetic mud formulations to maintain a stable emulsion of the internal brine phase. These materials include calcium and magnesium soaps of fatty acids and polyamines and amides and their mixtures (123,127). The specific chemistry of these additives depends on the nature of the continuous phase of the mud, ie, whether diesel oil, mineral oil, or a synthetic liquid. Lime is added along with the fatty acid to form the

soap emulsifier. Addition levels vary from supplier to supplier but are typically 6 to 22 kg m⁻³ (2–8 lb bbl) for both types of products.

Solids present in oil and synthetic muds must be kept wet with the nonaqueous phase to prevent coagulation and settling and mud instability. Oil-wetting agents are normally incorporated in the basic mud package. These materials are typically amines or quaternary ammonium salts having hydrocarbon chains of 10 or more carbon atoms. They also render clays or lignites oil-wet for use in viscosity and filtration control (128).

Defoamers (qv) are frequently needed for salty muds, although in very small quantities. A common defoamer is aluminum stearate [637-12-7]. Others include tributyl phosphate [55612-35-6], alkyl aryl sulfonates, silicones, and alcohols such as octanol [111-87-5].

Foaming agents maintain stable drilling foams in areas where minimal bottomhole pressures are required. A large number of chemicals generate drilling foams, including anionics, such as alkyl and alkylaryl sulfates and alkylarylsulfonates; nonionics, such as ethoxylated fatty alcohols and alkylaryl alcohols; and cationics, such as imidazolines and tertiary amines. The foaming agent must be chosen to handle a variety of possible contaminants (salt, crude oils, solids) and downhole temperatures (129,130).

Lost Circulation Control. To function properly, a drilling fluid must be circulated through the well and back to the surface. Occasionally, highly permeable or cavernous formations and fractured zones, both natural and induced by the mud pressure, are encountered and circulation is partially or completely lost. Loss of drilling fluid, owing to openings in the formation, can result in loss of hydrostatic pressure at the bottom of the hole and allow influx of formation fluids and possibly loss of well control. It is essential that circulation be regained for drilling to continue. A wide variety of materials can be added to the drilling fluid to seal off the lost circulation zones (131,132). The particle sizes of these materials are typically much larger than the particle sizes of solids normally suspended in the mud which are generally < 150 μm in diameter. Some of the same materials may be used for controlling filtration rates or for stabilizing shale formations but in such cases would be much smaller in particle size than a typical lost circulation material (LCM). Probably every material imaginable that can be pumped or moved down a well has been tried at one time or another; only lost circulation materials added to the drilling fluid and pumped down the drill string are discussed herein. Special soft- or hard-setting plugs and high filtration rate fluids containing diatomaceous earth (see DIATOMITE) which can be placed or injected into the loss zone also are available (132).

Lost circulation materials are flake, fiber, or granular-shaped particles. Each type is sold individually, often in two or more size grades, or two or more materials of different shapes may be sold as a blend (133–136). Materials of different shapes and sizes are often blended into the mud at the well site.

Some common flake-shaped LCMs consist of shredded cellophane and paper, mica (qv), rice hulls, cottonseed hulls, or laminated plastic. These materials lie flat across the opening to be sealed or are wedged into an opening such as a fracture. Some are sufficiently strong to withstand considerable differential pressure, whereas others are weak and the seal may be broken easily. Weaker flake materials typically are used near the surface or in combination with fibrous or granular additives.

Fibrous additives include a variety of cellulose fibers, sawdust, sugar-cane bagasse, paper, straw, leather (qv), and many others of similar size, shape, and availability. The larger fibers function by forming a brush-heap-type mat over the opening. The seal so formed may require smaller fibrous particles to stop seepage of mud through the mat. Fibers generally have little strength and cannot withstand high differential pressures. The brush-heap seal may extend far enough into the wellbore to be dislodged by the drill string, or be rigid enough to interfere with drill string movement.

Granular LCMs generally are much stronger than the other types and include ground rubber, nylon, plastics, limestone, gilsonite, asphalt, and ground nut shells, eg, walnut and pecan (see NUTS). Fine-, medium-, and coarse-size grades are available. Granular-shaped particles enter the opening, bridge it, and form a tight seal against further mud losses. Particle size and distribution are important for this mechanism to be effective. Bridging particles must generally have a diameter one-half the opening width for a fracture or one-third the diameter of a circular opening (137). An effective seal requires proper gradation of particle sizes (138). The advantage of this mechanism is that the seal is formed outside the wellbore and is not subject to drill string action.

Concentrations of LCMs are 14–143 kg m⁻³ (5–50 lb bbl). The higher concentrations are used in the form of a 10–20 m³ (60–120 bbl) pill that is placed across the loss zone until a seal is established that allows circulation to be regained.

Removal of Solids. Solids incorporated in the mud during drilling generally are separated mechanically, reduced by dilution, or removed chemically by flocculation. It is desirable to maintain a low concentration of drill solids (4–8 vol %) and in some cases total removal is required. In the latter case, the drilling fluid is clear, consisting of fresh water or brine, and high drilling rates can be achieved. Polymeric flocculants added in small (0.03–0.06 kg m⁻³ (0.01–0.02 lb bbl)) quantities maintain a clear liquid (139).

Polymers used for flocculating drill solids generally are high molecular weight polyacrylamides of varying degrees of hydrolysis (140). The polymer may be cationic, nonionic, or anionic, depending on the chemistry of the drilling fluid and the nature of the solids. At higher concentrations, some polymers act as protective colloids and stabilize and enhance the viscosity of bentonite suspensions and protect water-sensitive shales.

In low solids muds, vinyl acetate–maleic anhydride copolymers were once used to extend or enhance the viscosity of bentonite suspensions (141). This function is largely performed by polyacrylamides. The vinyl acetate–maleic anhydride copolymers can also have a flocculating effect on drill solids. Concentrations generally are quite low (0.14–0.57 kg m⁻³ (0.05–0.2 lb bbl)).

Lubricants and Spotting Fluids. The frictional resistance generated by the rotating drill string against the formation or casing may require extra torque if the hole is crooked or being drilled directionally. Considerable frictional resistance to raising and lowering the drill string may also occur; this is referred to as drag. Under certain conditions, such as in highly deviated holes, holes with frequent changes in direction, undergauge holes or poor drill string dynamics, or increased torque and drag entail a high risk of lost rig time, expensive pipe recovery operations, and limitation in well development (142). Excessive torque required to rotate the pipe generates greater strain on the drill pipe with the danger of twist-offs.

Another problem occurs when the drill pipe sticks to the wall of the borehole. This may be caused by running or pulling the pipe into an undergauge section of the hole, a key seat, or a bridge of cavings. Another form of sticking, known as differential pressure sticking, occurs when the drill pipe becomes motionless against a permeable formation and a portion of the contact area of the pipe is isolated by the filter cake. The differential pressure (mud pressure – formation pressure) acts on the isolated area to press the pipe against the wall.

To overcome these difficulties, drilling fluids are treated with a variety of mud lubricants available from various suppliers. They are mostly general-purpose, low toxicity, nonfluorescent types that are blends of several anionic or nonionic surfactants and products such as glycols and glycerols, fatty acid esters, synthetic hydrocarbons, and vegetable oil derivatives. Extreme pressure lubricants containing sulfurized or sulfonated derivatives of natural fatty acid products or petroleum-base hydrocarbons can be quite toxic to marine life and are rarely used for environmental reasons. Diesel and mineral oils were once used as lubricants at levels of 3 to 10 vol % but this practice has been curtailed significantly for environmental reasons.

Water-dispersible and sulfonated asphalt derivatives have been used as an aid to improve mud lubricity, although these are primarily for filtration control and shale protection. Solid lubricants, including graphite, plastic beads (143), glass beads (144), and ground walnut shells, may be incorporated to impart lubricity and reduce the risk of pipe sticking. Gilsonite and air-blown asphalts blended with a glycol-carrier are used to prevent pipe sticking and to act as lubricants for water-base muds (145,146).

Much confusion exists as to the best choice of lubricant additives for a given situation. Evaluation both in the laboratory and in the field is difficult because of the dynamic nature of the drilling fluid and the wide range of factors that influence drill string torque and drag. Liquid lubricants are used at concentrations of 0.25–4 vol %, solid materials at ca 6–29 kg m⁻³ (2–10 lb bbl).

If the drill string becomes differentially stuck, mechanical methods or spotting fluids can be applied, or the hydrostatic pressure can be reduced (147). In general, penetration of water- or oil-based spotting fluids into the interface between the filter cake and the pipe accompanied by dehydration and cracking results in reduction of differential pressure across the drill string (147,148). Spotting fluids are usually positioned in the open hole to completely cover the problem area.

Traditionally, various mixtures of diesel or mineral oil and commercial spotting fluid additives formulated with a density equal to that of the mud have been used to break the pressure lock. More recently a number of water-base spotting fluids have been developed which can be as effective as the oil-base systems (149). These fluids consist of a variety of surfactants, glycols (150,151), organic acids (152), glyceride esters (153), and vegetable oil derivatives. The newer water-base spotting fluids have the distinct advantage that they can be incorporated into the drilling fluid following use without

significantly impacting the toxicity of the system (154). Mud lubricants and spotting fluids, although not needed for every well, are essential for many deep directional wells. The consumption of these materials is difficult to estimate, and represents a relatively small fraction of the total drilling fluid additive market.

Corrosion Control. Drill string and casing corrosion can present serious problems in some drilling operations (see CORROSION AND CORROSION CONTROL). Corrosion is generally caused by oxygen dissolved or entrained in the mud as it is circulated through the well (155). Acid gases, such as carbon dioxide and hydrogen sulfide, contribute to corrosion, particularly hydrogen sulfide which can produce catastrophic drill string failures through stress cracking. Preventing entry of these corrosive gases into the well is the most effective method of control, but is not always feasible. Therefore additives are employed to counteract corrosive attack (156,157).

Oxygen corrosion is reduced significantly at pH > 11 and is generally lower at any alkaline pH than in an acidic environment. Maintaining a high pH is the most common means of corrosion control. Other methods include oxygen removal by scavengers such as sodium sulfite [7757-83-7] or ammonium bisulfite [17026-44-7] (157,158), or protecting the pipe from attack by coating with amines.

A high (>10) pH also affords protection against hydrogen sulfide and carbon dioxide. Zinc and iron compounds are used as sulfide scavengers (159). Zinc compounds include the carbonate, oxide, hydroxide, sulfate, and water-soluble organic chelates. High surface area iron oxides, particularly a synthetic magnetite, also are used (160). Scale inhibitors such as organic phosphonates and low molecular weight polyacrylates may be added to prevent buildup of carbonate scale on the pipe.

Sulfite concentrations for scavenging oxygen are typically in the 100– 300 mg /L range. Zinc compounds for sulfide scavenging are used at concentrations of 1.4–14 kg /m³ (0.5–5 lb / bbl); synthetic magnetite concentrations range to 140 kg /m³ (50 lb / bbl) and higher if the need arises.

Drilling Rate Enhancers. Drilling rates can be severely reduced when drilling formations that tend to stick to the surface of the drill bit, a situation known as bit balling. If the bit cannot be kept clean and free of sticky material, drilling performance suffers and the operational cost can increase substantially. The new generation synthetic polycrystalline diamond compact (PDC) bits are capable of achieving high penetration rates if the cutters are kept clean and the proper hydraulics and bit rotary speeds can be maintained (see TOOL MATERIALS). Use of an oil- or a synthetic-base mud is normally preferred when drilling with a PDC bit as there is much less tendency for bit balling to occur and high drilling rates can be realized (161). Water-base muds are more prone to bit balling and PDC bit performance can be significantly poorer than when a nonaqueous continuous-phase mud is used (162). Work is ongoing to overcome this performance difference. Additives being used to enhance the drilling performance of PDC bits in water-base muds include water-soluble glycols (163,164) and various terpenes (161). The use of a cationic mud has been reported to improve drilling rates as well (108,165).

Analytical and Test Methods

Procedures for determining drilling fluid properties are available (21,22,166). Tests and test methods are constantly reviewed by API committees to ensure acceptable accuracy when performed under field conditions as well as in the laboratory. API publications are republished as new tests are added or existing tests are modified.

Environmental Aspects

The disposal of waste drilling fluids and drill cuttings in the United States has long been regulated either by local authorities, the individual states, or by the federal government. These regulations continue to change. The offshore disposal of both diesel and mineral oil drilling fluids and associated cuttings has always been prohibited in U.S. waters. However discharge of mineral oil mud cuttings has been permitted in the North Sea and elsewhere as long as the oil content of the cuttings was below some regulatory limit. The regulatory oil-on-cuttings limit in some sectors of the North Sea is, as of this writing, being lowered significantly. There is a definite move toward alternative fluid systems, many of which are used in U.S. offshore areas.

The most significant change in the regulations on discharges of drilling fluids and cuttings in the Gulf of Mexico occurred in July 1986 when a new toxicity-based limit was placed on disposal (167). The specific toxicity limit is based on a 96-h bioassay test in which the drilling fluid is diluted with nine parts seawater and mixed well, the suspension is allowed to settle for one hour, and the suspended particulate phase (SPP) is decanted for testing. Mysid (*mysidopsis bahia*) shrimp are exposed to a range of SPP concentrations and the lethal concentration to 50% of the shrimp (LC₅₀) is calculated from the observed shrimp mortality at the end of 96 hours.

The toxicity limit is set at 30,000 ppm SPP. The LC₅₀ results must be reported monthly and at the end of the well. Any result below the 30,000 ppm SPP limit (the lower the LC₅₀ the higher the acute toxicity) may subject an operator to penalty. Other restrictions placed on offshore mud and cuttings discharges in 1986 and added since that time include a prohibition on discharge of any water-base mud containing diesel oil, diesel and mineral oil muds and cuttings, and muds using barite which contains more than 1 mg /kg mercury and 3 mg /kg cadmium. In addition the mud must pass a static sheen test. Regulatory limits are updated periodically, the most recent of which for the western Gulf of Mexico became effective in early 1994 (168).

Other regulations apply in different offshore drilling areas in the United States and around the world. All have had a profound effect on drilling fluid technology (169,170). Very few instances of water-base muds failing the mysid bioassay test exist in the 1990s. Operators and service companies have eliminated use of the more toxic additives, reformulated old mud systems, and developed new ones to ensure acceptable environmental performance based on pertinent regulations.

Economic Aspects

The price of drilling fluid additives varies according to company and location. Typical prices, including engineering service, for the more common additives from international drilling fluids companies having research laboratories and worldwide warehouse facilities are listed in Table 9. Price information on the full line of products normally is available from all drilling fluids companies. The materials listed in Table 9 are often sold at prices below those given depending on the nature of the contract between the supplier and operating companies. Estimates of the consumption of drilling fluid materials are available for chemicals used in the North Sea (171). Considerable changes have occurred in the makeup of drilling fluids in the 1980s and 1990s. Such changes are expected to continue.

Table 9. Prices of Drilling Fluid Additives

Function		Additive	Price, \$ /t ^a
increase density	barite		140
	hematite		155
increase viscosity	bentonite		165
	attapulgite		300
	xanthan gum		23,000
	hydroxyethylcellulose		11,500
reduce viscosity	sodium acid pyrophosphate		2,200
	lignite		
	mined		510
	causticized		1,200
	chrome lignosulfonate		1,150
	chrome-free lignosulfonate		1,500
reduce filtration rate	sodium polyacrylate (liquid)		6,500
	corn starch		1,300

shale stability	modified starch	3,000
	carboxymethyl cellulose (PAC)	11,000
	sodium polyacrylate	6,500
	modified lignite	4,000
	polyacrylamide	
viscosity	powder	13,000
	liquid	7,100
	sodium hydroxide	1,050
alkalinity control	lime	220
	potassium hydroxide	2,200
lost circulation control	walnut shells	750
	mica	590
	cellulose fibers	1,500
	sodium chloride	240
miscellaneous inorganic compounds	potassium chloride	550
	calcium chloride	910
	gypsum	225
	sodium carbonate	500
	sodium bicarbonate	650

^a Texas Gulf Coast, 1994.

BIBLIOGRAPHY

"Drilling Fluids" in *ECT* 1st ed., Suppl. 1, pp. 246–258, by G. R. Gray and E. E. Huebotter, Baroid Div., National Lead Co.; in *ECT* 2nd ed., Vol. 7, pp. 287–307, by E. E. Huebotter and G. R. Gray, Baroid Div., National Lead Co.; "Drilling Fluids" under "Petroleum" in *ECT* 3rd ed., Vol. 17, pp. 143–167, by R. K. Clark and J. J. Nahm, Shell Development Co.

1. N. J. Adams and T. Charrier, *Drilling Engineering, A Complete Well Planning Approach*, PennWell Books, Tulsa, Okla., 1985.
2. P. L. Moore, *Drilling Practices Manual*, PennWell Books, 2nd ed., Tulsa, Okla., 1985.
3. A. T. Bourgoyne, K. K. Millheim, M. E. Chenevert, and F. S. Young, *Applied Drilling Engineering*, Society of Petroleum Engineers, Richardson, Tex., 1991.
4. G. R. Gray and H. C. H. Darley, *Composition and Properties of Oil Well Drilling Fluids*, 5th ed., Gulf Publishing Co., Houston, Tex., 1988.
5. **U.S. Pats. 1,575,944 and 1,575,945** (Mar. 9, 1926), B. K. Stroud (to National Pigments and Chemical Co.).
6. *World Oil*, 51 (June 1994).
7. A. H. Hale and F. K. Mody, *Society of Petroleum Engineers Symposium on Oilfield Chemistry*, SPE 25180, New Orleans, La., Mar. 2–5, 1993.
8. A. H. Hale and A. K. Dewan, *SPE Drilling Eng.* **5**, 109 (1990).
9. R. K. Clark, R. F. Scheuerman, H. Rath, and H. van Laar, *J. Pet. Tech.* **28**, 719 (1976).
10. R. P. Steiger, *J. Pet. Tech.* **34**, 1661 (1982).
11. R. B. Bennett, *J. Pet. Tech.* **36**, 975 (1984).
12. L. J. Fraser, *SPE Drilling Eng.* **7**, 20 (1992).
13. J. P. Simpson, *J. Pet. Tech.* **31**, 643 (1979).
14. F. B. Growcock, S. L. Andrews, and T. P. Frederick, *Society of Petroleum Engineers Drilling Conference*, IADC SPE Paper 27450, Dallas, Tex., Feb. 15–18, 1994.
15. **U.S. Pat. 5,252,554** (Oct. 12, 1993), H. Mueller and co-workers, (to Henkel Kommanditgesellschaft auf Atkien and Baroid Ltd.).
16. P. Kenney, *Oil Gas J.*, 88 (Nov. 8, 1993).
17. **U.S. Pat. 5,096,883** (Mar. 17, 1992), J. D. Mercer and L. L. Nesbit (to Union Oil Co. of Calif.).
18. S. Park, D. Callum, and A. D. McLean, *Society of Petroleum Engineers Annual Meeting*, SPE 26354, Houston, Tex., Oct. 3–6, 1993.
19. E. Gao and A. C. Young, *Society of Petroleum Engineers Drilling Conference*, SPE IADC 29425, Amsterdam, the Netherlands, Feb. 28–Mar. 2, 1995.
20. J. D. A. McKee, K. Dowrick, and S. J. Astleford, in Ref. 19, SPE IADC 29405.
21. *Standard Procedure for Field Testing Water-Based Drilling Fluids*, API RP 13B-1, 1st ed., American Petroleum Institute, Washington, D.C., June 1990.
22. *Standard Procedure for Field Testing Oil-Based Drilling Fluids*, API RP 13B-2, 2nd ed., American Petroleum Institute, Washington, D.C., Dec. 1991.
23. *Specification for Drilling Fluid Materials*, API Spec. 13A, 15th ed., Sect. 2, American Petroleum Institute, Washington, D.C., May 1, 1993.
24. Ref. 23, Sect. 3.
25. N. E. Blomberg, B. Melberg, A. Boe, E. A. Jacobsen, and S. Aarrestad, *J. Pet. Tech.* **35**, 969 (1984).
26. S. W. Golis, *Society of Petroleum Engineers Annual Meeting*, SPE 13156, Houston, Tex., Sept. 16–19, 1984.
27. J. P. Sloan, J. P. Brooks, and S. F. Dear, *J. Pet. Tech.* **27**, 15 (1975).
28. T. Mondshine, *Oil Gas J.*, 79 (Jan. 19, 1981).
29. J. M. McLarty, J. W. Dobson, and M. A. Dick, *Offshore Technology Conference*, OTC 7352, Houston, Tex., May 3–6, 1993.
30. J. D. Downs, in Ref. 7, SPE 25177.
31. H. van Olphen, *An Introduction to Clay Colloid Chemistry*, 2nd ed., John Wiley & Sons, Inc., New York, 1977.
32. Ref. 23, Sect. 4, 5, and 6.
33. **U.S. Pat. 2,702,788** (Feb. 22, 1955), L. R. Dawson (to Milwhite Co.).
34. J. L. Lummus and L. J. Field, *Pet. Engr.* **59** (Mar. 1968).
35. J. L. Burba, W. E. Holman, and C. R. Crabb, *Society of Petroleum Engineers Drilling Conference*, IADC SPE 17198, Dallas, Tex., Feb. 28–Mar. 2, 1988.
36. W. F. Bradley, *Am. Mineral.* **25**, 405 (1940).
37. L. L. Carney and R. L. Meyer, *Society of Petroleum Engineers Annual Meeting*, SPE Paper 6025, New Orleans, La., Oct. 1976.
38. L. L. Carney and N. Guven, *Soc. Pet. Eng. J.* **20**, 385 (1980).
39. Ref. 23, Sect. 7 and 8.
40. J. W. Jordan, *J. Phys. Chem.* **53**, 294 (1949).
41. J. W. Jordan, B. J. Hook, and C. M. Finlayson, *J. Phys. Chem.* **54**, 1196 (1950).
42. **U.S. Pat. 4,450,095** (May 22, 1984), C. M. Finlayson (to NL Industries).
43. T. L. Hughes, T. G. J. Jones, and O. H. Houwen, *SPE Drilling Complet.* **8**, 157 (1993).
44. K. S. Kang and D. J. Pettitt, in R. L. Whistler and J. N. BeMiller, eds., *Industrial Gums*, 3rd ed., Academic Press, Inc., New York, 1993, Chapt. 20, pp. 537–578.
45. Ref. 23, Sect. 10.
46. A. J. Desmarais and R. F. Wint, in Ref. 44, Chapt. 19, pp. 505–535.
47. **U.S. Pat. 3,953,335** (Apr. 27, 1976), J. M. Jackson.

48. K. S. Kang and D. J. Pettitt, in Ref. 44, Chapt. 13, pp. 341–397.

49. R. D. Carico, *Society of Petroleum Engineers Regional Meeting*, SPE Paper 5870, Long Beach, Calif., Apr. 8–9, 1976.

50. G. P. Sanz, M. C. Gunningham, H. C. Lau, and A. J. Samuel, *SPE Drilling Complet.* **9**, 139 (1994).

51. J. W. Powell, C. F. Parks, and J. M. Scheult, *Society of Petroleum Engineers Arctic Technology Conference*, SPE 22066, Anchorage, Alaska, May 29–31, 1991.

52. H. Maier, M. Anderson, C. Karl, K. Magnisuson, and R. L. Whistler, in Ref. 44, Chapt. 8, pp. 181–226.

53. A. G. Kadaster, G. J. Guild, G. L. Hanni, and D. D. Schmidt, *SPE Drilling Eng.* **44**, 191 (1992).

54. R. C. Portnoy, R. D. Lundberg, and E. R. Werdein, *IADC/SPE Drilling Conference*, IADC SPE 14795, Dallas, Tex., Feb. 10–12, 1986.

55. **U.S. Pat. 4,447,338** (May 8, 1984), R. D. Lundberg, T. O. Walker, C. P. O'Farrell, and H. S. Makowski (to Exxon Research and Engineering Co.).

56. M. R. Rabaioli, F. Miano, T. P. Lockhart, and G. Burrafato, in Ref. 7, SPE 25179.

57. G. R. Gray and H. C. H. Darley, *Composition and Properties of Oil Well Drilling Fluids*, 5th ed., Gulf Publishing Co., Houston, Tex., 1988, pp. 582–583.

58. **U.S. Pat. 3,479,287** (Nov. 18, 1969), J. C. Floyd and F. J. Shell (to Phillips Petroleum Co.).

59. C. M. Frost, J. J. Hoepner, and W. W. Fowkes, *J. Chem. Eng. Data* **4**, 173 (1959).

60. W. W. Fowkes and C. M. Frost, *U.S. Bur. Mines Rep. Invest.*, 5611 (1960).

61. F. E. Brauns and D. A. Brauns, *The Chemistry of Lignin*, Suppl. Vol., Academic Press, Inc., New York, 1960, pp. 313–386.

62. L. L. Carney, *American Petroleum Institute Southern District Meeting*, API Paper 926-15-M, Division of Production, Houston, Tex., Mar. 4–6, 1970.

63. L. S. Park, *Society of Petroleum Engineers International Symposium on Oilfield Chemistry*, SPE 16281, San Antonio, Tex., Feb. 4–6, 1987.

64. R. D. Wilcox and M. A. Jarrett, *Society of Petroleum Engineers Drilling Conference*, IADC SPE 17201, Dallas, Tex., Feb. 28–Mar. 2, 1988.

65. B. G. Chesser and D. P. Enright, *Society of Petroleum Engineers Annual Conference*, SPE 8224, Las Vegas, Nev., Sept. 23–26, 1979.

66. **U.S. Pat. 4,518,510** (May 21, 1985), P. A. Gleason and I. E. Brase (to National Starch and Chemical Corp.).

67. **U.S. Pat. 3,998,742** (Dec. 21, 1976), T. O. Walker (to Texaco, Inc.).

68. T. O. Walker, H. L. Dearing, and J. P. Simpson, *World Oil*, 93 (Nov. 1983).

69. J. A. Neal, J. M. L. Zdybak, and W. G. Bannerman, *Oil Gas J.*, 118 (Mar. 25, 1985).

70. T. C. Mondshine and J. D. Kercheville, *Baroid News Bull.* **15**, 12 (1963).

71. G. R. Gray, J. L. Foster, and T. S. Chapman, *Trans. AIME* **146**, 117 (1942).

72. J. N. BeMiller, in Ref. 44, Chapt. 21, pp. 579–600.

73. Ref. 23, Sect. 11.

74. T. K. Haack, D. A. Shaw, and D. E. Greenley, *Oil Gas J.*, 81 (Jan. 6, 1986).

75. D. C. Thomas, *Soc. Pet. Engrs. J.* **22**, 171 (1982).

76. **U.S. Pat. 3,256,115** (June 14, 1966), R. O. Stearns and D. J. Weintritt (to National Lead Co.).

77. **U.S. Pat. 3,294,681** (Dec. 27, 1966), R. O. Stearns and D. J. Weintritt (to National Lead Co.).

78. T. C. Mondshine, *Oil Gas J.*, 84 (Mar. 21, 1966).

79. **U.S. Pat. 3,956,141** (May 11, 1976), T. O. Walker (to Texaco, Inc.).

80. J. R. Pettigrove, *Drilling*, 120 (May 15, 1964).

81. S. Palumbo, D. Giacca, M. Ferrari, and P. Pirovano, *Society of Petroleum Engineers International Symposium on Oilfield Chemistry*, SPE 18477, Houston, Tex., Feb. 8–10, 1989.

82. Ref. 23, Sect. 9.

83. C. S. Scanley, *World Oil*, 122 (July 1959).

84. **U.S. Pat. 3,441,504** (Apr. 29, 1969), W. C. Browning (to Milchem, Inc.).

85. **U.S. Pat. 3,766,229** (Oct. 16, 1973), F. Turner and J. W. Nahm (to Dresser Industries, Inc.).

86. **U.S. Pat. 3,956,140** (May 11, 1976), J. W. Nahm and D. A. Rowe (to Dresser Industries, Inc.).

87. W. R. Clements, M. A. Jarrett, and E. K. Morton, *Society of Petroleum Engineers Annual Conference*, SPE 16686, Dallas, Tex., Sept. 27–30, 1987.

88. J. P. Plank, *Oil Gas J.*, 40 (Mar. 2, 1992).

89. M. Hille, *Society of Petroleum Engineers International Symposium on Oilfield and Geothermal Chemistry*, SPE 13558, Phoenix, Ariz., Apr. 9–11, 1985.

90. A. J. Son, T. M. Ballard, and R. E. Loftin, *SPE Drilling Eng.* **2**, 209 (1987).

91. **U.S. Pat. 4,547,299** (Oct. 15, 1985), J. M. Lucas (to Milchem Inc.).

92. K. H. W. Ujma and J. P. Plank, *SPE Drilling Eng.* **4**, 41 (1989).

93. T. Hemphill, *World Oil*, 81 (Nov. 1989).

94. L. Altazin, *Oil Gas J.*, 76 (Aug. 4, 1980).

95. T. C. Mondshine, *Oil Gas J.*, 70 (July 14, 1969).

96. M. E. Chenevert, *J. Pet. Tech.* **22**, 1309 (1970).

97. A. H. Hale, F. K. Mody, and D. P. Salisbury, *SPE Drilling Complet.* **8**, 207 (1993).

98. J. Elward-Berry and E. W. Thomas, *Society of Petroleum Engineers Drilling Conference*, IADC SPE 27453, Dallas, Tex., Feb. 15–18, 1994.

99. G. Radenti, S. Palumbo, and G. Zucca, *Pet. Engr. Int.*, 32 (Sept. 1987).

100. P. I. Reid, R. C. Minton, and A. Twynam, *Society of Petroleum Engineers European Petroleum Conference*, SPE 24979, Cannes, France, Nov. 16–18, 1992.

101. K. E. Gillenwater and C. R. Ray, *Oil Gas J.*, 99 (Mar. 20, 1989).

102. F. J. Shell, *Drilling-DCW*, 47 (Nov. 1969).

103. D. P. Vesely, *Society of Petroleum Engineers Rocky Mountain Regional Meeting*, SPE 17511, Casper, Wyo., May 11–13, 1988.

104. D. E. O'Brien and M. E. Chenevert, *J. Pet. Tech.* **25**, 189 (1973).

105. D. B. Anderson and C. D. Edwards, *Pet. Eng.*, 105 (Sept. 1977).

106. **U.S. Pat. 5,208,216** (May 4, 1993), C. D. Williamson, D. W. Fong, and R. K. Gabel (to Nalco Chemical Co.).

107. S. E. Alford, *Society of Petroleum Engineers Drilling Conference*, SPE IADC 21936, Amsterdam, Mar. 11–14, 1991.

108. T. Hemphill, R. Valenziano, P. Bale, and B. Sketchler, *Oil Gas J.*, 60 (June 8, 1992).

109. O. Welch and L. Lee, *Oil Gas J.*, 53 (July 13, 1992).

110. T. W. Beihoffer, D. S. Dorrough, C. K. Deem, and D. D. Schmidt, *Oil Gas J.*, 47 (Mar. 16, 1992).

111. T. W. Beihoffer and co-workers, *Society of Petroleum Engineers Annual Conference*, SPE 24588, Washington, D.C., Oct. 4–7, 1992.

112. R. Bland, *Oil Gas J.*, 54 (June 29, 1992).

113. P. I. Reid, G. P. Elliott, R. C. Minton, and D. A. Burt, *Society of Petroleum Engineers/EPA Exploration & Production Environmental Conference*, SPE 25989, San Antonio, Tex., Mar. 7–10, 1993.

114. J. D. Downs, E. van Oort, D. I. Redman, D. Ripley, and B. Rothmann, *Society of Petroleum Engineers Offshore European Conference*, SPE 26699, Aberdeen, Scotland, Sept. 7–10, 1993.

115. A. J. Twynam, P. A. Caldwell, and K. Meads, in Ref. 98, IADC SPE 27451.

116. **U.S. Pat. 4,780,220** (Oct. 25, 1988), T. E. Peterson (to Hydra Fluids, Inc.).

117. M. E. Chenevert, *Oil Gas J.*, 60 (July 17, 1989).

118. **U.S. Pat. 5,076,373** (Dec. 31, 1991), A. H. Hale, G. C. Blytas, and A. K. R. Dewan (to Shell Oil Co.).

119. **U. S. Pat. 5,149,690** (Sept. 22, 1992), A. D. Patel and H. C. McLaurine (to M-I Drilling Fluids Co.).

120. J. R. Bruton and H. C. McLaurine, in Ref. 18, SPE 26327.

121. D. E. Clark and S. Benaissa, *Society of Petroleum Engineers Asia Pacific Oil & Gas Conference*, SPE 25321, Singapore, Feb. 8–10, 1993.

122. J. P. Simpson, T. O. Walker, and G. Z. Jiang, in Ref. 98, IADC SPE 27496.

123. J. A. Headley, T. O. Walker, and R. W. Jenkins, *Society of Petroleum Engineers Drilling Conference*, SPE IADC 29404, Amsterdam, Feb. 28–Mar. 2, 1995.

124. W. S. Cagle and L. F. Schwertner, *Oil Gas J.*, 61 (Mar. 27, 1972).

125. N. Davis and C. E. Tooman, *SPE Drilling Eng.* **4**, 47 (1989).

126. R. F. Burdyn and L. D. Wiener, *World Oil*, 101 (Nov. 1957).

127. D. K. Clapper and D. P. Salisbury, *World Oil*, 125 (Nov. 1984).

128. F. B. Growcock, C. F. Ellis, and D. D. Schmidt, *SPE Drilling Complet.* **9**, 39 (1994).

129. W. J. McDonald, L. J. Remont, W. A. Rehm, and M. E. Chenevert, *Development of Drilling Foams for Geothermal Applications*, SAND79-7122, Sandia National Laboratories, Albuquerque, N.M., Jan. 1980.

130. P. B. Rand and O. J. Montoya, *Aqueous Foam Surfactants for Geothermal Drilling Fluids: I. Screening*, SAND79-2036, Sandia Laboratories, Albuquerque, N.M., Jan. 1980.

131. G. R. Gray and H. C. H. Darley, *Composition and Properties of Oil Well Drilling Fluids*, 5th ed., Gulf Publishing Co., Houston, Tex., 1988, pp. 595–598.

132. J. U. Messinger, *Last Circulation*, PennWell Books, Tulsa, Okla., 1981.

133. G. E. Loeppke, D. A. Glowka, and E. K. Wright, *J. Pet. Tech.* **42**, 328 (1990).

134. T. M. Nayberg, *SPE Drilling Eng.* **2**, 229 (1987).

135. **U.S. Pat. 5,207,282** (May 4, 1993), G. F. Fuh, N. Morita, D. L. Whitfill, and D. A. Strah (to Conoco Inc.).

136. A. Ali, C. L. Kalloo, and B. Singh, *SPE Drilling Complet.* **9**, 32 (1994).

137. C. J. Coberly, *Drilling and Production Practices*, American Petroleum Institute, New York, 1937, pp. 189–201.

138. C. Gatlin and C. E. Nemir, *J. Pet. Tech.* **13**, 575 (1961).

139. J. P. Gallus, J. L. Lummus, and J. F. Fox, *Trans. AIME* **213**, 70 (1958).

140. **U.S. Pat. 3,637,031** (Jan. 25, 1972), J. D. Hull and R. E. Finch (to Standard Brands Chemical Industries, Inc.).

141. D. B. Anderson, *Oil Gas J.*, 65 (Mar. 26, 1973).

142. J. E. Fontenot, *Oil Gas J.*, 58 (Jan. 14, 1974).

143. **U.S. Pat. 4,063,603** (Dec. 20, 1977), J. J. Rayborn.

144. **U.S. Pat. 4,123,367** (Oct. 31, 1978), A. A. Dodd.

145. **U.S. Pat. 5,114,598** (May 19, 1992), J. J. Rayborn and J. P. Dickerson (to Sun Drilling Products Corp.).

146. R. G. Bland, W. T. Micho, and H. J. Howard, *Society of Petroleum Engineers Annual Conference*, SPE 24591, Washington, D.C., Oct. 4–7, 1992.

147. N. Adams, *Pet. Eng.*, 40 (Oct. 1977).

148. H. D. Outmans, *Trans. AIME* **213**, 265 (1958).

149. R. K. Clark and S. G. Almquist, *SPE Drilling Eng.* **7**, 121 (1992).

150. **U.S. Pat. 4,614,235** (Sept. 30, 1986), I. G. Keener, J. C. Newlove, R. C. Portnoy, D. F. Fehler, and K. W. Pober (to Exxon Chemical Patents, Inc.).

151. **U.S. Pat. 5,057,234** (Oct. 15, 1991), R. G. Bland and D. K. Clapper (to Baker Hughes Inc.).

152. **U.S. Pat. 5,247,992** (Sept. 28, 1993), R. Lockhart.

153. **U.S. Pat. 5,002,672** (Mar. 26, 1991), J. R. Hayes and G. J. Campbell (to Turbo-Chem. International, Inc.).

154. W. S. Halliday and D. K. Clapper, *Society of Petroleum Engineers Drilling Conference*, SPE IADC 18684, New Orleans, La., Feb. 28–Mar. 3, 1989.

155. C. C. Patton, *Oil Gas J.*, 66 (July 22, 1974); *Oil Gas J.*, 163 (July 29, 1974); *Oil Gas J.*, 76 (Aug. 5, 1974).

156. H. E. Bush, *Society of Petroleum Engineers Annual Meeting*, SPE Paper 5123, Houston, Tex., Oct. 6–9, 1974.

157. T. Cox, *Drilling Contractor*, 37 (May–June 1975).

158. **U.S. Pat. 3,899,293** (Aug. 12, 1975), H. E. Bush (to NL Industries, Inc.).

159. R. L. Garrett, R. K. Clark, L. L. Carney, and C. K. Grantham, *J. Pet. Tech.* **31**, 787 (1979).

160. **U.S. Pat. 4,008,775** (Feb. 22, 1977), I. Fox (to Ironite Products Co.).

161. R. K. Clark and T. Hemphill, *SPE Drilling Complet.* **9**, 176 (1994).

162. J. M. Simon, *Society of Petroleum Engineers Annual Conference*, SPE 24602, Washington, D.C., Oct. 4–7, 1992.

163. D. P. Enright, W. M. Dye, and F. M. Smith, *SPE Drilling Eng.* **7**, 15 (1992).

164. **U.S. Pat. 5,007,489** (Apr. 16, 1991), D. P. Enright, W. M. Dye, F. M. Smith, and A. C. Perricone (to Baker Hughes Inc.).

165. J. R. Hasley, K. E. Dunn, W. R. Reinhardt, T. S. Carter, and W. M. Duncan, in Ref. 98, IADC SPE 27539.

166. *Standard Procedure for Laboratory Testing Drilling Fluids*, API RP 13I, 4th ed., American Petroleum Institute, Washington, D.C., June 1990.

167. *U.S. Fed. Reg.* **51**(131), 24897 (July 9, 1986).

168. *U.S. Fed. Reg.* **58**(231), 63964 (Dec. 3, 1993).

169. R. Bleier, A. J. J. Leuterman, and C. Stark, *J. Pet. Tech.* **45**, 6 (1993).

170. R. K. Clark, *J. Pet. Tech.* **46**, 804 (1994).

171. C. M. Hudgins, *J. Pet. Tech.* **46**, 67 (1994).

R. K. Clark
Shell E&P Technology Company

ENHANCED OIL RECOVERY

Enhanced oil recovery (EOR) requires the successful application of chemical, chemical engineering, and petroleum engineering technologies (1). EOR accounted for about 3.2% of the world's 1994 oil production, ca $3 \times 10^5 \text{ m}^3 \text{ d}$ ($1.9 \times 10^6 \text{ bbl d}$) (2). In 1994, U.S. EOR production ($113,000 \text{ m}^3 \text{ d}$ ($709,000 \text{ bbl d}$)) represented about 10% of total production. As the petroleum industry becomes more dependent on increasing production from existing fields, the use of EOR is expected to grow.

Primary and secondary oil recovery together recover only 25–50% of the oil originally in place in a reservoir. For example, in 89 years of production, the Glenn Pool Field (northeast Oklahoma) produced only 20% of the original oil in place. After conventional oil recovery operations, more than 56 billion m^3 ($350 \times 10^9 \text{ bbl}$) of oil are expected to remain in known U.S. reservoirs (3). The most likely location of principal new oil fields is Alaska and deep water offshore. The increasing costs of discovering and developing these reserves make unrecovered oil in known fields an economically attractive target in part because much of the infrastructure is also already in place.

Oil price decline since 1981 has resulted in significant changes in the EOR technology being developed and field tested. Injection of steam (qv) or oil-miscible gases remains of great interest. Micellar polymer flooding can efficiently recover oil and was the focus of a large research and development effort from 1970 to 1986, but this process is prohibitively expensive at 1994 oil prices. The concentration of surfactant is often 2–5% by weight; the water-thickening polymer concentration is 100–1000 ppm.

The focus of more recent work has been the use of relatively low concentrations of additives in other oil recovery processes. Of particular interest is the use of surfactants (qv) as CO_2 (4) and steam mobility control agents (foam). Combinations of older EOR processes such as surfactant-enhanced alkaline flooding and alkaline–surfactant–polymer flooding show promise of improved cost effectiveness.

The Nature of Oil Reservoirs

Oil reservoirs are layers of porous sandstone or carbonate rock, usually sedimentary. Impermeable rock layers, usually shales, and faults trap the oil in the reservoir. The oil exists in microscopic pores in rock. Various gases and water also occupy rock pores and are often in contact with the oil. These pores are interconnected with a complicated network of microscopic flow channels. The weight of overlying rock layers places these fluids under pressure. When a well penetrates the rock formation, this pressure drives the fluids into the wellbore. The flow channel size, wettability of flow channel rock surfaces, oil viscosity, and other properties of the crude oil determine the rate of this primary oil production.

As reservoir pressure is reduced by oil production, additional recovery mechanisms may operate. One such mechanism is natural water drive. Water from an adjacent more highly pressured formation is forced into the oil-bearing formation by the pressure differential between the formations. Another mechanism is gas drive. Expansion of a gas cap above the oil as oil pressure declines can also drive additional oil to the wellbore. Produced gas may be reinjected to maintain gas cap pressure as is done on the Alaskan North Slope. Additional oil may also be produced by compaction of the reservoir rock as oil production reduces reservoir pressure.

As the natural pressures in the reservoir decrease, oil production declines. The oil well may then be placed on-pump to maintain production at economic levels. The pump draws oil to the surface and lowers the height of the fluid column in the wellbore. The pressure of a column of fluid can decrease the rate of fluid entry into the wellbore.

Primary production typically recovers 10–25% of the oil originally in the reservoir. Efficiency of primary production is related to oil properties, reservoir properties, geometric placement of oil wells, and the drilling and completion technology used to drill the wells and prepare them for production. Pumping the well can maintain production at economic levels for years.

Waterflooding. Injection wells are used when the natural pressures driving fluids to production wells are depleted and pumping is no longer economical. Fluid injection repressurizes the reservoir, restoring a driving force and promoting oil production. For economic reasons, water is the usual injection fluid. Water injection or waterflooding is usually termed secondary oil recovery. It accounts for about 40% of the total U.S. oil production. Additional oil recovery by waterflooding is typically 15–25% of the oil originally in the reservoir.

Determining and using the optimum pattern and arrangement of production and injection wells can have a significant effect on oil recovery and production rates (5). Infill drilling and horizontal production wells can drain oil reservoirs more efficiently (6). Horizontal injection wells have also improved oil recovery by increasing volumetric sweep efficiency and increasing fluid injection rates (7). One 600–1200-m long horizontal well can replace several vertical wells decreasing both overall drilling and operational costs (8).

Oil Recovery Mechanisms

There are two principal mechanisms of enhanced oil recovery: increasing volumetric sweep efficiency of the injected fluid and increasing oil displacement efficiency by the injected fluid. In both, chemicals are used to modify the properties of an injected fluid whether water, steam, a miscible gas such as CO_2 or natural gas, or an immiscible gas, usually nitrogen. Poor reservoir volumetric sweep efficiency is the greatest obstacle to increasing oil recovery (9).

Wettability is defined as the tendency of one fluid to spread on or adhere to a solid surface (rock) in the presence of other immiscible fluids (5). As many as 50% of all sandstone reservoirs and 80% of all carbonate reservoirs are oil-wet (10). Strongly water-wet reservoirs are quite rare (11). Rock wettability can affect fluid injection rates, flow patterns of fluids within the reservoir, and oil displacement efficiency (11). Rock wettability can strongly affect its relative permeability to water and oil (5,12). When rock is water-wet, water occupies most of the small flow channels and is in contact with most of the rock surfaces as a film. Crude oil does the same in oil-wet rock. Alteration of rock wettability by adsorption of polar materials, such as surfactants and corrosion inhibitors, or by the deposition of polar crude oil components (13), can strongly alter the behavior of the rock (12).

When water is injected into a water-wet reservoir, oil is displaced ahead of the injected fluid. Injection water preferentially invades the small- and medium-sized flow channels or pores. As the water front passes, unrecovered oil is left in the form of spherical, unconnected droplets in the center of pores or globules of oil extending through interconnected rock pores. In both cases, the oil is completely surrounded by water and is immobile. There is little oil production after injection water breakthrough at the production well (5).

In an oil-wet rock, water resides in the larger pores, oil exists in the smaller pores or as a film on flow channel surfaces. Injected water preferentially flows through the larger pores and only slowly invades the smaller flow channels resulting in a higher produced water:oil ratio and a lower oil production rate than in the water-wet case.

Injection Well Considerations. Fluid injection rate can have a significant effect on oil recovery economics. Flow is radial from the wellbore into the reservoir. Thus the region near the injection wellbore acts as a choke for the entire reservoir.

Addition of surfactant to the injection water (14,15) can displace the oil remaining near the well. The lower oil saturation results in an increase in the water relative permeability (5). Therefore, a greater water injection rate may be maintained at a given injection pressure. Whereas ultimate oil recovery may not be increased, the higher water injection rate can increase oil production rates improving oil recovery economics. Alternatively, a lower injection pressure can be used. Thus smaller and cheaper injection pumps may be used to maintain a given injection rate. The concentration of surfactant in the injection

water is relatively high (1–3%). However, the total amount of surfactant used is not great because it is necessary only to displace the oil from a 2–3-m radius around the injection well (see SURFACTANTS).

Decreased injection rates resulting from formation damage, ie, reduction of the rock fluid carrying capacity, near injection wells can reduce oil production rates at offset (adjacent) production wells. Formation damage may result from invasion of rock capillaries by solid particles in wellbore fluids during well drilling and completion. Plugging of rock capillaries adjacent to fractures by fine solid particles in fracturing fluids may also occur. Acidizing the rock immediately adjacent to the wellbore can dissolve clays (qv), silica particles, and precipitates plugging rock flow channels.

Precipitate formation can occur upon contact of injection water ions and counterions in formation fluids. Solids initially present in the injection fluid, bacterial corrosion products, and corrosion products from metal surfaces in the injection system can all reduce near-wellbore permeability. Injectivity may also be reduced by bacterial slime that can grow on polymer deposits left in the wellbore and adjacent rock. Strong oxidizing agents such as hydrogen peroxide, sodium perborate, and occasionally sodium hypochlorite can be used to remove these bacterial deposits (16–18).

Formation damage can also be caused by chemical and physical interactions of fluid and rock. Low salinity injection fluids are often preferred to obtain maximum viscosity from a given amount of water-soluble polymer. However, low salinity fluids can cause swelling of water-expandable clays. This swelling reduces the fluid-carrying capacity of rock flow channels. Because clays act as the cementing medium in many sandstone formations, this swelling weakens cementation and results in the release of mineral fine particles which can migrate to constrictions and plug the flow channels.

Long-lasting stabilization of water-swelling clays may be achieved by using materials such as hydroxyaluminum (19) and certain quaternary ammonium salt polymers (20). These clay stabilizers are polymers containing many cationic sites. Quaternary ammonium salt polymers have been used in drilling fluids, completion fluids, acidizing treatments, and hydraulic fracturing as well as in injection water and injected steam for enhanced oil recovery (20). A single molecule adsorbs onto minerals at multiple sites. Simultaneous ion exchange (qv) at these many adsorption sites is required for clay stabilizer desorption. Adsorption is long lasting and limited chemical treatment volumes are needed. Treatment concentration is usually 0.1–1.0%. The total amount of polymer injected is dependent on the well treatment design and rock properties. For injection water and injected steam, it is usually necessary to treat a 2–3-m radius from the injection wellbore. The addition of potassium hydroxide to injection waters has also been used to stabilize clays and maintain injection rates (21).

Injection Fluids. Whereas water is the most commonly used injection fluid, other fluids can provide higher oil recovery efficiency. Injecting gases miscible with reservoir crude oil can result in low interfacial tension promoting a high oil displacement efficiency (22). The process of miscible gas flooding using carbon dioxide (qv) is depicted in Figure 1. Other suitable gases include natural gas and flue gas. Carbon dioxide is of most interest in the United States; hydrocarbon miscible projects represent 80% of Canadian EOR production (2). Injection of hydrocarbon miscible gas is used in the Alaskan North Slope and is under study for use in the North Sea.

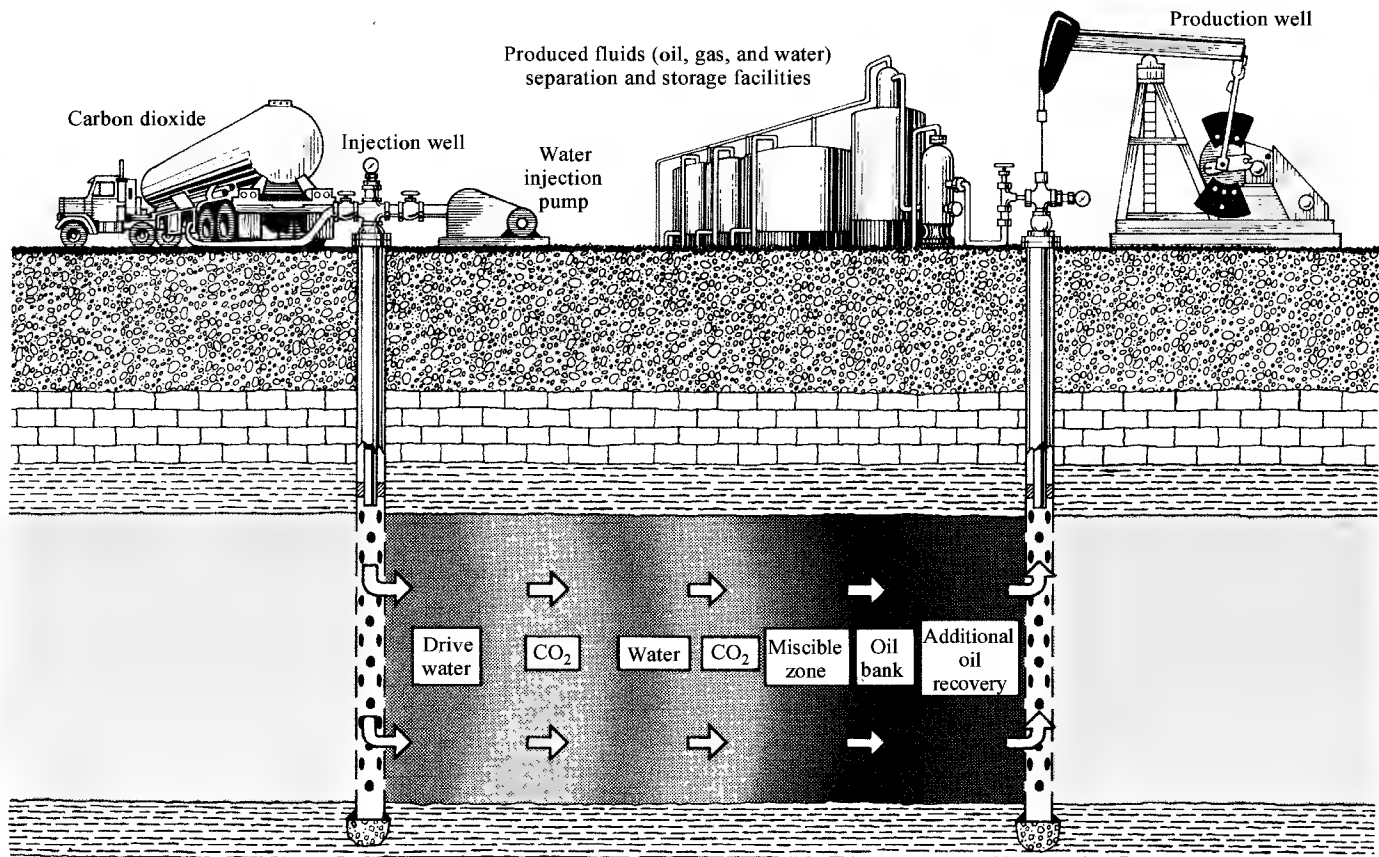


Fig. 1. Carbon dioxide flooding. The WAG process, in which a CO₂ slug is followed by alternate water and CO₂ injections, is usually employed. The viscosity of the oil is reduced providing more efficient miscible displacement.

Courtesy of the U.S. Dept. of Energy.

Most CO₂ miscible EOR projects are located in the west Texas Permian Basin where as much as two-thirds of the oil remains after waterflooding. An incremental (10%) recovery is typical for Permian Basin CO₂ floods, which could correspond to as much as $0.5 \times 10^9 \text{ m}^3$ ($3\text{--}4 \times 10^8 \text{ bbl}$) (23). Pipelines (qv) connect these large EOR projects to natural CO₂ sources in Colorado and New Mexico. Industrial point sources of CO₂ have also been used for projects in other areas.

The pressure composition requirement for miscibility limits the oil reservoirs in which this technology has been applied. However, the low injected fluid viscosity often results in poor volumetric sweep efficiency.

Supercritical CO₂ (22,24) and various hydrocarbon injectants (22,25) undergo physical interactions with crude oil that result in stripping out of the low molecular weight components, which increases oil production (see SUPERCRITICAL FLUIDS). The rapid or gradual development of miscibility with the remaining crude oil constituents results in oil mobilization. Either partial or complete miscibility with the oil may be developed depending on the nature of the injectant, crude oil properties, and reservoir conditions, particularly temperature. However, interaction of the injectant with the crude oil can alter rock wettability and thus reduce injection rates and decrease oil recovery.

Another method of using CO₂ is called cyclic CO₂ stimulation or huff'n'puff (26). A limited amount of CO₂ is injected into a reservoir over hours or days. The well is then shut in for a soak period of days to weeks to allow the CO₂ to interact with the crude oil, swelling the oil and reducing its viscosity.

The well is then opened, the CO₂ provides a solution gas drive, and oil mobilized by the CO₂ soak is produced.

Nonmiscible gases such as nitrogen have been used as EOR injection fluids. Oil recovery mechanisms include volatilization of low molecular weight components of the crude oil and displacement of oil from the top of the reservoir (27). The latter mechanism occurs as a result of gravity override of the low density injectant. That is, the fluid migrates to the upper part of the formation, resulting in channeling of injected fluid through the upper portion of the reservoir. A low volumetric sweep efficiency results.

Gas injection into a gas cap overlying an oil reservoir is considered an EOR method. The resulting repressurization of the reservoir promotes additional oil production. Reinjection of natural gas is responsible for a significant fraction of Alaskan North Slope oil production.

High temperature steam (qv) is also used for recovery of viscous crude oils (28). Heat from the steam thins the oil, reducing viscosity and increasing mobility. The mobilized oil is produced at offset production wells. In heavy oil fields, water flooding is often omitted and steam injection begun immediately after primary production. Steam injection temperature is typically 175–230°C in California oil fields. Injection temperature can reach 300°C in Canadian and Venezuelan EOR projects.

The injection of large volumes of steam, steam flooding, is used to mobilize oil which is produced at offset production wells. Smaller volumes of steam are injected in the cyclic steam stimulation or huff'n'puff process (Fig. 2). Many wells are placed on several cycles of steam stimulation and then used as injection or production wells in steam flood projects.

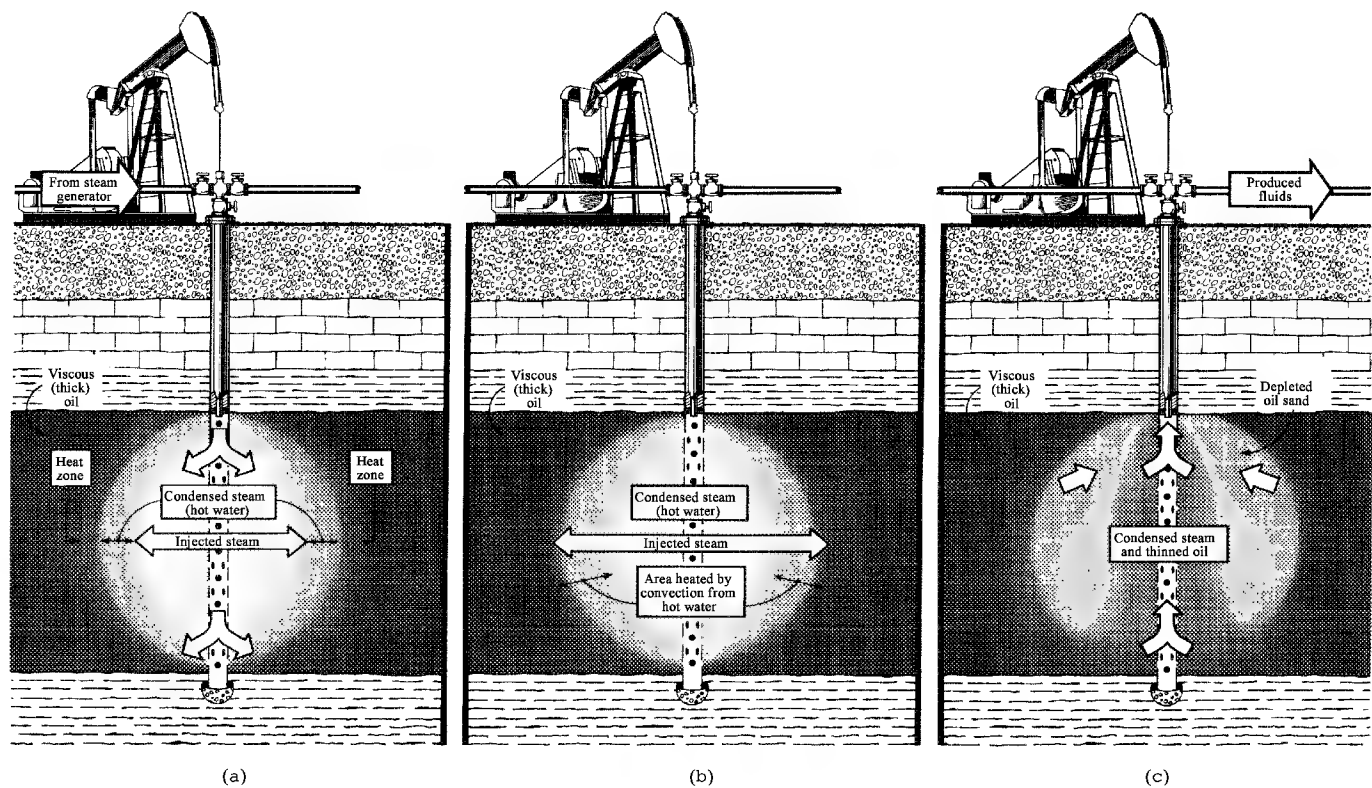


Fig. 2. Cyclic steam stimulation of an oil well: (a) steam, injected into a well over a period of days or weeks in a heavy oil reservoir, introduces heat (huff) that, coupled with (b), alternate soak periods lasting a few days to allow (c) a production phase of weeks or months (puff), thins the oil. This process may be repeated until production falls below a profitable level.

Courtesy of the U.S. Dept. of Energy.

Improving Volumetric Sweep Efficiency. Volumetric sweep efficiency is determined by the permeability and wettability distribution in the reservoir and by the properties of injected fluids. High permeability rock streaks or layers (thief zones) and natural or induced rock fractures can channel the injected fluid through a small portion of the reservoir, resulting in a low rock volumetric sweep efficiency. Low viscosity injection fluids exhibit poor volumetric sweep efficiency and lead to low oil production. Thus, proper diagnosis of the cause of poor volumetric sweep efficiency is critical in designing a successful well treatment. For example, the sealing fractures process requires different well treatment designs than reducing the permeability of thief zones.

Both sodium silicate gelation (29) and *in situ* cross-linking of organic polymers (30,31) can reduce the permeability of fractures and high permeability streaks. Polymers are usually injected at concentrations of 1000–5000 ppm. *In situ* cross-linking treatments are restricted to fractures and the near-wellbore region owing to the kinetics of the cross-linking process. The polymer may be injected into the well with a cross-linker or the cross-linker may be injected after the polymer. The well is shut in for from 1–7 days to allow cross-linking to occur depending on the treatment. Normal injection operations are then resumed.

The most commonly used polymers are partially hydrolyzed polyacrylamides (32). The optimum degree of hydrolysis depends on the application, injection water composition, and reservoir conditions (33,34). More salt-tolerant acrylamide copolymers may permit this technology in higher salinity injection water (35). Field applications of cross-linked xanthan gum have also been reported (36).

Chromium(III) Cr(III), compounds have largely replaced Al(III) compounds as cross-linkers (36–39). Cr(III) acetate [1066-30-4] cross-links acrylamide polymers rapidly. Cr(III) complexes composed of strong ligands such as glycolate or malonate give extended polyacrylamide gelation times compared to salts such as Cr(III) acetate (40). Delaying cross-linking permits the use of this technology at higher (up to 150°C) reservoir temperatures. Sodium bisulfite and thiourea have been used to reduce injected Cr(VI) to the reactive Cr(III) species that promotes cross-linking (41). Gradual dissolution of colloidal Cr(OH)₃ can also delay cross-linking (42) as can Cr(III) propionate (43). Injection of unhydrolyzed polyacrylamide followed by *in situ* hydrolysis also delays cross-linking (44).

Sequential injection of partially hydrolyzed polyacrylamide and cross-linker is giving way to treatments in which the two components are injected together. Large-volume injection well treatments have been carried out using a partially hydrolyzed polyacrylamide (mol wt 5–20 × 10⁶) at 3000–8500 ppm and a chromium triacetate cross-linker. Using colloidal dispersion gels, the polymer is cross-linked with aluminum(III) citrate before injection well treatment. Successful results have been reported for reducing injection water channeling in fractured reservoirs (45). When it is desirable to have rapid cross-linking take place, blends of chromium triacetate and hydrochloric acid have been used (46). Applied shear can also affect gelation time (47).

Organic cross-linkers, which include glyoxal (48) and formaldehyde (qv), have also been used. Use of hypohalite salts (49) and epichlorohydrin (50) promotes gel stability. Phenol-formaldehyde cross-linking systems have been used to produce stable acrylamide copolymer gels at temperatures above 75°C and brine hardness levels above 2000 ppm (51).

Copolymers of sodium acrylate with sodium 2-acrylamido-2-methylpropane sulfonate (AMPS) or *N,N*-dimethylacrylamide (52) have been used to prepare cross-linked systems at high temperatures and salinity. Chromium cross-linked gels, prepared from a 3:1 blend of partially hydrolyzed

polyacrylamide and guar gum, are said to have higher strength and stability than gels prepared from partially hydrolyzed polyacrylamide alone (53).

Cross-linked xanthan gums have also been used to reduce the permeability of thief zones. Trivalent chromium is the preferred cross-linker (54). Cross-linker effectiveness is less at high salinity. However, Cr(III) has been used in the field at salinities as great as 166,000 ppm total dissolved solids (55).

Proper placement of the treatment fluid in the reservoir is critical to treatment success. Careful sizing of the treatment and choice of injection rates are required because overtreatment can cause plugging of the oil-containing rock and excessive reduction of the injection rate. Even after a well-designed treatment, the fluid injection rate is often significantly less than before well treatment. Many successful applications of this technology in waterfloods and in surfactant polymer floods have been reported. Wells in CO₂ EOR projects have also been treated using this technology.

Polymerization may also occur *in situ*. Reactive monomers such as acrylamide in concentrations of 2–5 wt % and various additives including a free-radical polymerization initiator may be used (56,57). A difunctional monomer such as N,N'-methylenebis(acrylamide) can be added to the formulation to form a cross-linked polymer *in situ*. Low viscosity aqueous monomer solutions can be injected at higher rates. Low viscosity monomer solutions preferentially enter high permeability zones to a greater extent than do non-Newtonian polymer solutions (58). If no difunctional monomer is used, the viscous polymer mass may be slowly dissolved by injection water increasing water viscosity and providing a second means of increasing oil recovery.

Lignosulfonates may be cross-linked *in situ* using Cr(III) (59) or an acidic gas such as CO₂ (60). Cross-linked lignosulfonate can be an effective plugging agent at high formation temperatures. Lignosulfonate concentration is usually 2–3 wt %. This gelation technology has been evaluated in field tests in both waterflood and steam injection wells. Blends of lignosulfonate and sodium silicate have also been used (61). Other systems cross-linked *in situ* for water or steam injection wells are phenol–formaldehyde (62), urea–formaldehyde (63), furfuryl alcohol (64), formaldehyde resin plus sulfonated tannin extract (65), and formaldehyde resin and alkali kraft lignin (66).

Surfactant precipitation may be used for in-depth permeability reduction of thief zones (67). Thief zones have a low oil saturation owing to the preferential flow of injected fluids through high permeability rock. This process is based on the sequential injection of a slowly propagating ionic surfactant followed by an aqueous spacer containing no surfactant. Then a more rapidly propagating ionic surfactant of the opposite charge type is injected. The oppositely charged surfactants gradually mix in the high permeability portions of the reservoir (thief zones) causing precipitation. The precipitates plug flow channels. The cumulative effect is to reduce permeability in the most flooded portions of the reservoir, diverting injectant to rock zones containing higher oil saturations. Whereas these rock zones originally had a lower permeability than the thief zones, owing to the plugging of the thief zones, these other rock zones become the preferred flow path for injected fluid. The result is increased oil production from the less flooded zones. The economically limiting factors in the use of this process are cost and low propagation rate of the cationic surfactant.

The chemistry of rock surfaces can affect volumetric treatment effectiveness and economics. Cationic chemicals such as metal ion cross-linkers and cationic polymers can adsorb on mineral surfaces, particularly clays, by ion-exchange (qv) processes. The subsequently lower concentration of materials in solution can decrease treatment effectiveness by reducing the rate of polymer cross-linking and decreasing gel strength. Polyethylene glycol ethers can form a high viscosity emulsion on contact with residual crude oil to plug thief zones (68).

Both *in situ* cross-linking of partially hydrolyzed polyacrylamides (69) and injection of quaternary ammonium salt polymers having long hydrophilic side chains (70) have been used to reduce the permeability of water-producing zones adjacent to production wells. This permeability reduction decreases the produced water:oil ratio as does injection of polyacrylamide in high hardness brine to reduce permeability (71).

The polymers exist in saline solution as tightly coiled chains and are readily adsorbed owing to relatively low solubility in hard water. Subsequent injection of soft, low salinity water uncoils the adsorbed polymer chains increasing water viscosity and reducing rock permeability. This technology could also be used to reduce the permeability of thief zones adjacent to injection wells. However, mechanical isolation of these zones may be necessary for cost-effective treatments.

Polymer Flooding. Even in the absence of fractures and thief zones, the volumetric sweep efficiency of injected fluids can be quite low. The poor volumetric sweep efficiency exhibited in waterfloods is related to the mobility ratio, M , the mobility of the injected water in the highly flooded (low oil saturation) rock, m_w , divided by the mobility of the oil in oil-bearing portions of the reservoir, m_o (72,73). The mobility ratio is related to the rock permeability to oil, k_{ro} , and injected water, k_{rw} , and to the viscosity of these fluids by the following equation:

$$M = \frac{m_w}{m_o} = \frac{(k_{rw} \eta_o)}{(k_{ro} \eta_w)}$$

The terms η_w and η_o represent the viscosity of the aqueous and oil phases, respectively.

The polymer flooding process is depicted in Figure 3. The displacing or driving fluid may be steam, supercritical carbon dioxide, hydrocarbon miscible gases, nitrogen, or solutions of surfactants or polymers instead of water. The volumetric sweep efficiency increases with lower mobility ratio values (74). A mobility ratio of 1.0 or lower is considered optimum. The mobility of water is often high relative to that of oil. Steam and oil-miscible gases such as supercritical carbon dioxide exhibit even higher mobility ratios. Consequently, these more expensive injectants can have low volumetric sweep efficiencies.

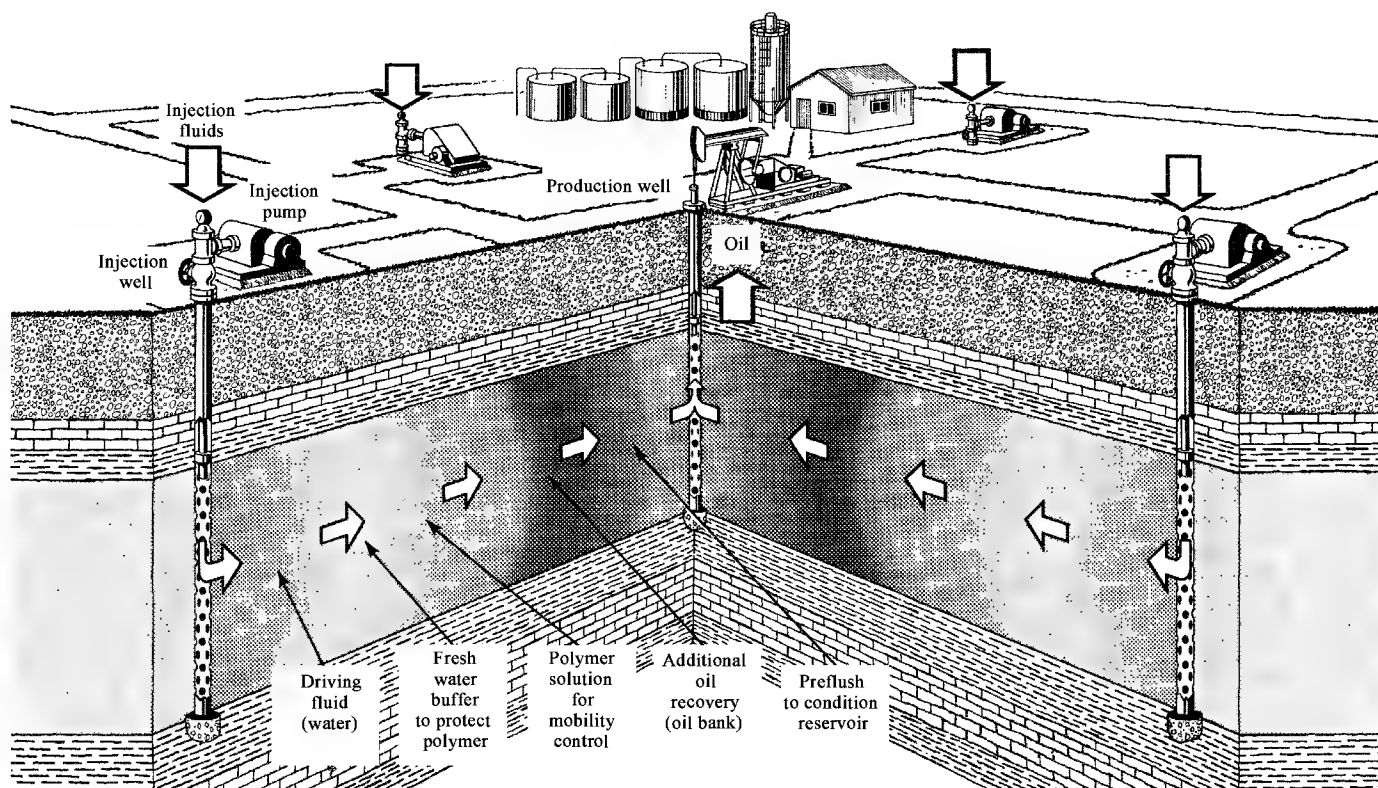


Fig. 3. This polymer flooding method requires a preflush to condition the reservoir, the injection of a polymer solution for mobility control to minimize

channeling, and a driving fluid (water) to move the polymer solution and resulting oil bank to production wells.

Courtesy of the U.S. Dept. of Energy.

Mobility control agents reduce the mobility ratio. The early mobility control agents were partially hydrolyzed polyacrylamides having molecular weights of $1\text{--}5 \times 10^6$ and xanthan gum, a biopolymer (75). Virtually all field projects have used polymers from one of these two classes. Variations in polymer molecular weight and structure have been made to improve performance properties. Relatively low (100 ppm for fresh water, 1000 ppm or more for saline systems) polymer concentrations can significantly increase injected water viscosity. Adsorption of these polymers on rock can result in a decrease in rock permeability to aqueous fluids (residual resistance). This permeability reduction persists during long periods of water injection. Some polymer field projects have exhibited injected water permeability reductions, attributed to residual resistance effects, that have lasted for more than three years and seven years, respectively, after polymer injection (76).

Each EOR polymer type has important advantages and significant disadvantages (Table 1). When dissolved in more saline waters, xanthan gum produces a higher apparent viscosity than the same concentration of partially hydrolyzed polyacrylamide (78). Xanthan gum is more soluble in saline waters than are polyacrylamides, particularly in injection waters containing divalent metal ions. Xanthan gum also generally adsorbs less on rock surfaces and is substantially more resistant to shear degradation than polyacrylamides (77). However, xanthan gum is also more expensive and the extensional viscosity of the semirigid xanthan molecule is less than that of the flexible polyacrylamide (79). Both polymers cross-link easily in the presence of transition metals.

Table 1. Properties of EOR Polymers ^a

Property	Polyacrylamide	Xanthan gum
brine tolerance	very limited, especially to Ca^{2+} , Mg^{2+}	good to both mono- and divalent cations
shear stability	undergoes irreversible shear degradation	reversible shear thinning
maximum use temperature, °C	71–82 ^b	71–77
hydrolytic stability	hydrolysis promoted by acid or base; partially hydrolyzed product more sensitive to Ca^{2+} , Mg^{2+}	hydrolytic depolymerization promoted by acid or base especially at high temperatures
oxidative stability	susceptible	particularly susceptible especially at high temperatures
microbial degradation	susceptible	very susceptible

^a Ref. 77.
^b In very low salinity reservoirs temperatures can go to 107–121°C.

In addition to the normal problems of completely dissolving particles of water-thickening polymers, xanthan gum contains insoluble residues which decrease polymer injectivity. Various methods of reducing insolubles content and improving xanthan solution injectivity are available (80–87). None appears economically viable. Oxygen scavengers (88) and bactericides (77,89) are commonly used to stabilize injected polyacrylamide and xanthan gum solutions (90–102).

At high polymer concentrations, pyruvate ring-opened xanthan exhibits increased tolerance to divalent metal ions in high density completion fluids (103). However, at low polymer concentrations, xanthan containing the intact pyruvate ring exhibits higher brine solution viscosity and better filterability than its ring-opened analogue (104). A xanthan gum containing pyruvate rings in most of the polymer repeat units has been produced by a proprietary strain of *Xanthamonas campestris* (105). Xanthamonas bacteria have also been used to produce a polymer, having a glucose and mannose unit in a 2:1 ratio, which is claimed to be a better water viscosifier than xanthan gum (106). Another promising microbial polysaccharide is scleroglucan (107) which has been evaluated for North Sea applications. The high temperature behavior of different microbial polysaccharides (qv) has been studied (108). In highly saline media, another biopolymer, succinoglycan, more readily flows through microporous media than does xanthan gum (108). This glycan forms higher viscosity solutions than does xanthan gum at equal concentrations.

Most polyacrylamides used as mobility control agents are partially hydrolyzed or are acrylamide–acrylic acid (or sodium acrylate) copolymers produced by emulsion copolymerization (109). Emulsion polymers are used to avoid high shear degradation and undissolved solid particle problems. Another method of avoiding these problems is acrylamide solution polymerization at the wellhead. The polymerization can be designed to proceed at adequate rates and in saline injection waters to provide polymers of adequate viscosity characteristics (110). Polyacrylamide is usually hydrolyzed in base to produce a random distribution of acrylate groups (111). Acid hydrolysis results in a more block-like distribution of acrylate units (112). Electrostatic repulsion of the anionic carboxylate groups elongates the polymer chain of partially hydrolyzed polyacrylamides, increasing the hydrodynamic volume and solution viscosity. Cobalt-60, ^{60}Co , irradiation has been used to initiate polymerization and prepare particularly high molecular weight polyacrylamides (113).

Maximum freshwater viscosity of polyacrylamide occurs at ca 35% hydrolysis; maximum viscosity in a Ca_2+ -containing brine occurs at 10–15% hydrolysis (114). Metal ions interact with carboxylate groups reducing their mutual repulsion and thus decreasing hydrodynamic volume and solution viscosity. Divalent metal ions reduce viscosity more than monovalent ones (115). The principal mode of polyacrylamide decomposition at elevated temperature in the absence of oxygen is hydrolysis (116,117). Thus, the concentration of divalent metal ions has an effect on viscosity retention at high temperature. Chelating and sequestering agents have been proposed to reduce the adverse effect of divalent (114) and multivalent metal ions on polyacrylamide solution viscosity (118,119). Proper well completion, particularly perforation design, reduces polyacrylamide shear degradation during injection (120).

Acrylamide copolymers designed to reduce undesired amide group hydrolysis, increase thermal stability, and improve solubility in saline media have been studied for EOR applications (121–128). These polymers still tend to be shear sensitive. Most copolymers evaluated for EOR have been random copolymers. However, block copolymers of acrylamide and AMPS also have utility (129).

Excessive hydrolysis of polyacrylamide *in situ* can promote undesirable polymer precipitation in the reservoir. The rate of this hydrolysis decreases with increasing level of anionic comonomers such as AMPS (130).

Acrylamide graft copolymers such as those with starch (qv)(131), dextran (132), and lignin (qv) (133), have been studied to try to reduce copolymer costs. A general disadvantage of acrylamide copolymers is greater cost compared to partially hydrolyzed polyacrylamides.

Polymer Interactions. Various methods of utilizing polymer interactions to modify solution viscosity are under study. Polymer association complexes (134–138), which substantially increase water viscosity at quite low polymer concentrations, offer potentially improved cost effectiveness compared to the flood polymers. However, use has not progressed beyond the laboratory testing stage. Polymer association complexes form micelle-like structures by interaction of hydrophobic groups. Polymers forming such structures include nonylphenoxy polyethylene glycol acrylates (139), acrylamide terpolymers containing hydrophobic alkylacrylamides (140–142), and poly(styrene-*co*-maleic anhydride) vinylbenzylpolyglycol ethers (143).

The substantial decrease of polyacrylamide solution viscosity in mildly saline waters can be utilized to increase injection rates. A quaternary ammonium salt polymer can be added to the polyacrylamide solution to function as a salt and reduce solution viscosity (144). If the cationic charge is in the polymer backbone and substantially shielded from the polyacrylamide by steric hindrance, formation of an insoluble interpolymer complex can be delayed long enough to complete polyacrylamide injection. Upon contacting formation surfaces, the quaternary ammonium salt polymer is adsorbed reducing

solution salinity and increasing aqueous fluid viscosity away from the wellbore. Cationic polymer adsorption also prevents undesirable precipitation of a polymer complex. By using a clay stabilizing quaternary ammonium salt polymer, formation damage associated with low salinity polyacrylamide solvents can be reduced (145).

Propagation of enhanced oil recovery chemicals through rock is critical to the success of an EOR project. Mechanical entrapment of polymers as well as adsorption can reduce the effective polymer concentration (146,147). Calcium ions in formation and injection waters increase anionic polymer adsorption. Three mechanisms have been proposed. The first is reduced electrostatic repulsion between anionic polymer and negatively charged mineral surfaces. The second, specific interaction of calcium ions with the polymer, reduces polymer solubility. The third mechanism is fixation of calcium ions on mineral surfaces, reducing the net negative surface charge and generating new sites for polymer adsorption (148).

Sacrificial adsorption agents such as lignosulfonates (148–151) can be used to reduce the adsorption of more expensive polymers and surfactants. Other chemicals tested include poly(vinyl alcohol) (152), sulfonated poly(vinyl alcohol) (153), sulfonated poly(vinylpyrrolidinone) (153), low molecular weight polyacrylates (154), and sodium carbonate (155).

Surfactants for Mobility Control. Water, which can have a mobility up to 10 times that of oil, has been used to decrease the mobility of gases and supercritical CO₂ (mobility on the order of 50 times that of oil) used in miscible flooding. Gas:oil mobility ratios, *M*, can be calculated by the following (22):

$$M = \frac{[(k_g / \mu_s) + (k_w / \mu_w)] }{ [(k_o / \mu_o) + (k_w / \mu_w)] }$$

where *k* refers to permeability, *μ* to viscosity, and the subscripts *g*, *s*, *o*, and *w* to gas, miscible solvent, oil, and water, respectively. The water may be injected simultaneously with the gas or in alternate slugs with the gas (WAG process). X-ray computerized tomography of core floods has demonstrated the increased volumetric sweep efficiency attained in the WAG process (156) compared to injection of CO₂ alone. The design parameters most affecting WAG CO₂ flood oil recovery are CO₂ and water slug sizes, produced gas:oil ratio as a function of time, and total volume of injected CO₂ (157) (see Fig. 1).

The WAG process has been used extensively in the field, particularly in supercritical CO₂ injection, with considerable success (22,157,158). However, a method to further reduce the viscosity of injected gas or supercritical fluid is desired. One means of increasing the viscosity of CO₂ is through the use of supercritical CO₂-soluble polymers and other additives (159). The use of surfactants to form low mobility foams or supercritical CO₂ dispersions within the formation has received more attention (160–162). Foam has also been used to reduce mobility of hydrocarbon gases and nitrogen. The behavior of foam in porous media has been the subject of extensive study (4). X-ray computerized tomographic analysis of core floods indicate that addition of 500 ppm of an alcohol ethoxyglycerylsulfonate increased volumetric sweep efficiency substantially over that obtained in a WAG process (156).

One reason for widespread interest in the use of surfactants as gas mobility control agents is the effectiveness at concentrations of <0.1 wt % (156,163). Some surfactants are effective below their critical micelle concentration (164). This low chemical requirement can significantly improve process economics.

Among the classes of surfactants studied for this application are alcohol ethoxylates and their sulfate and sulfonate (157,165–168) and carboxylate (169) derivatives, alkylphenol ethoxylates (170), alpha-olefin sulfonates (169), and alkylated diphenylether disulfonates (171). Increased linear carbon chain length, decreased branching, and increased ethoxy group chain length increase foam stability (165). By using mixtures of surfactants or surfactants plus an alcohol having the same carbon chain length (172), foam stability, injected breakthrough time at the core outlet, and oil recovery are maximized. Use of organosilicon polymers as steam-foaming agents is said to reduce the formation damage caused by steam-promoted silica dissolution (173). Addition of a water-thickening polymer to the aqueous phase may stabilize the foam (174).

In addition to the mobility control characteristics of surfactants, critical issues in gas mobility control processes are surfactant salinity tolerance, hydrolytic stability under reservoir conditions, surfactant propagation through the reservoir, and foam stability in the presence of crude oil saturations. Lignosulfonate has been reported to increase foam stability and function as a sacrificial adsorption agent (175). Addition of sodium carbonate or sodium bicarbonate to the surfactant solution reduces surfactant adsorption by increasing the aqueous-phase pH (176).

Alcohol ethoxysulfates have been used in field tests as nitrogen (177) and carbon dioxide (178) foaming agents. Field use of alcohol ethoxysulfates is restricted to low temperature formations owing to its limited hydrolytic stability at low pH and elevated temperature (179). It has been reported that some foams can reduce residual oil saturation, not by oil displacement, but by emulsification and imbibition of the oil into the foam (180).

Gravity override of low density steam leads to poor volumetric sweep efficiency and low oil recovery in steam floods. Nonchemical methods of improving steam volumetric sweep efficiency include completing the injection well so steam is only injected in the lower part of the oil-bearing zone (181), alternating the injection of water and steam (182), and horizontal steam injection wells (183,184). Surfactants frequently are used as steam mobility control agents to reduce gravity override (185). Field-proven surfactants include C_{11–18} alpha-olefin sulfonates (AOS), alkyltoluene sulfonates, and neutralized dimerized alpha-olefin sulfonic acid. A review of 17 field projects is available (162).

Addition of long-chain (C_{8–20}) alcohols to AOS or alkylaromatic sulfonates increases foam strength and permits the use of lower surfactant concentrations (186). Increasing the carbon number in alpha-olefin sulfonates to >25 increases foam strength (187,188). In alkylaromatic sulfonates, longer linear alkyl groups (189,190) or dialkyl substitution (191) has the same effect. Other alkylaromatic sulfonates containing benzene, toluene, or xylene rings (190,192), two fused aromatic groups (193), and the diarylether group (194) have been favorably evaluated as steam-foaming agents. The neutralized dimer of an alpha-olefin sulfonate has also been used (195).

High temperature steam cools and eventually condenses as it propagates through the oil reservoir. To maintain foam strength as the steam cools, a noncondensable gas, usually nitrogen or methane, is often added to the injectant composition (196). Methods of calculating the optimum amount of noncondensable gas to use are available (197).

Critical parameters affecting surfactant performance are surfactant propagation rate and surfactant stability at steam temperatures that can reach more than 316°C. Surfactant propagation rate can be reduced by adsorption, precipitation, and partitioning into the oil phase. Anionic surfactant adsorption increases with increasing salinity and decreases with increasing temperature (198). A numerical model of foaming agent transport has been developed (199).

Additives can improve surfactant propagation. Both anionic surfactant partitioning and precipitation increase with increasing calcium ion concentration (200) so minimizing divalent metal ion concentration in the surfactant solution is desirable. Injection of a surfactant preslug containing NaCl converts clays from the calcium to the sodium form and reduces later ion-exchange processes that add Ca²⁺ ions in the surfactant solution (201,202). The use of a hydrotrope such as sodium xylene sulfonate has been reported to increase oil recovery in laboratory steam-foam flood tests (203). Hydrotropes are additives that increase surfactant solubility. They also may function as sacrificial adsorption agents or act as foam stabilization agents.

Steam-foaming agents that efficiently mobilize heavy crude oil by heat transfer can reduce the residual oil saturation. This can increase foam stability and improve the diversion of subsequently injected steam into oil saturated zones thereby increasing oil recovery (204).

Thermal stability of the foaming agent in the presence of high temperature steam is essential. Alkylaromatic sulfonates possess superior chemical stability at elevated temperatures (205,206). However, alpha-olefin sulfonates have sufficient chemical stability to justify their use at steam temperatures characteristic of most U.S. steamflood operations. Decomposition is a desulfonation process which is first order in both surfactant and acid concentrations (206). Because acid is generated in the decomposition, the process is autocatalytic. However, reservoir rock has a substantial buffering effect.

The addition of high pH agents such as sodium hydroxide to the surfactant solution has been reported to increase foam strength and stability (207). Sodium hydroxide, sodium carbonate, and trona (Na₂CO₃–NaHCO₃) form precipitates with calcium ions to improve surfactant propagation (208,209). These additives can also maintain the pH at a high enough value to reduce the rate of surfactant decomposition. In addition, the added base may interact with organic acids naturally found in the crude oil. The resulting soap generation provides surfactant to more efficiently displace oil (207). The consequent lower oil saturation can result in a more stable foam.

Water-soluble polymers (qv) can increase the viscosity of the foam external phase. This improves foam stability and reduces mobility. Gelation of

the foam external phase can reduce chemical requirements to plug thief zones and fractures (210).

Improving Oil Displacement Efficiency. The use of relatively large (ca 2–5 wt %) concentrations of surfactants to increase oil displacement efficiency has been studied extensively (27,211,212). This method, called the micellar flooding or surfactant–polymer flooding, usually involves the injection of a brine preflush to adjust reservoir salinity. The preflush is followed by injection of a micellar slug comprised of the surfactant, a cosurfactant (usually a C_{4–} alcohol), and a hydrocarbon. A polymer solution is then injected to reduce viscous fingering of the drive fluid into and through the micellar slug. Viscous fingering causes dilution of the surfactant, reduced contact of the micellar slug with the crude oil, and trapping of some of the micellar slug in the reservoir. These effects reduce oil recovery. A freshwater buffer to protect the polymer follows, prior to addition of the driving fluid, ie, water, to move the chemicals and the resulting oil bank to the well.

Process effectiveness depends on maintaining an ultralow (ca 10^{–10} N m (10^{–3} dynes cm)) interfacial tension between the injected surfactant slug and the crude oil (213). The effect of petroleum composition on oil solubilization by surfactants has been the subject of extensive study (214).

By 1980, research and development shifted from relatively inexpensive surfactants such as petroleum sulfonates to more costly but more effective surfactants tailored to reservoir and crude oil properties. Critical surfactant issues are performance in saline injection waters, adsorption on reservoir rock, partitioning into reservoir crude oil, chemical stability in the reservoir, interactions with the mobility control polymer, and production problems caused by resultant emulsions. Reservoir heterogeneity can also greatly reduce process effectiveness. The decline in oil prices in the early 1980s halted much of the work because of the relatively high cost of micellar processes.

In the 1990s, the thrust of surfactant flooding work has been to develop surfactants which provide low interfacial tensions in saline media, particularly seawater; require less cosurfactant; are effective at low concentrations; and exhibit lower adsorption on rock. Nonionic surfactants such as alcohol ethoxylates, alkylphenol ethoxylates (215) and propoxylates (216), and alcohol propoxylates (216) have been evaluated for this application. More recently, anionic surfactants have been used (216–230).

The alpha-olefin sulfonates (AOS) have been found to possess good salt tolerance and chemical stability at elevated temperatures. AOS surfactants exhibit good oil solubilization and low interfacial tension over a wide range of temperatures (219,231), whereas less salt tolerant alkylaromatic sulfonates exhibit excellent chemical stability. The nature of the alkyl group, the aryl group, and the aromatic ring isomer distribution can be adjusted to improve surfactant performance under a given set of reservoir conditions (232,233).

Cosurfactant requirements can be minimized using a surfactant having a short-branched hydrophobe or a branched-alkyl substituent on an aromatic group (232,234) and a long ethoxy group chain (234). Blends of surfactants optimized for seawater or reservoir brine salinity include linear alkylxylene sulfonate–alcohol ether sulfate mixtures (235).

High (1–10⁰ %) concentrations of lignosulfonate have sufficient interfacial activity to increase oil recovery from unconsolidated sands (236). Lignosulfonates and petroleum sulfonates undergo a synergistic interaction resulting in ultralow interfacial tension and substantially increased oil recovery (237). Low molecular weight ethoxylated, sulfated, or sulfonated lignin phenols used alone in surfactant floods recover >75% of the oil present (238). Alkylated oxidized lignins have also been used (239). Paper (qv) industry spent sulfite liquors function in a similar manner if these do not contain Ca²⁺ or Mg²⁺. These divalent metal ions cause petroleum sulfonate precipitation (240).

The effect of temperature, pressure, and oil composition on oil recovery efficiency have all been the subjects of intensive study (241). Surfactant propagation is a critical factor in determining the EOR process economics (242). Surfactant retention owing to partitioning into residual crude oil can be significant compared to adsorption and reduce surfactant propagation rate appreciably (243).

Various low cost sacrificial agents decrease surfactant adsorption on reservoir rock and increase the surfactant propagation rate. These agents include lignosulfonates and chemically modified lignosulfonates (4,75,151). Sodium saccharide wastes from wood pulping (244) and low molecular weight polyethylene oxide (245) have also been used. Alkaline chemicals (208,209), particularly sodium silicate (246), which precipitate in the presence of divalent metal ions, can increase the surfactant propagation rate. Internixing of polymer mobility control fluid with a previously injected surfactant slug can result in surfactant–polymer interactions affecting interfacial behavior and reducing oil displacement efficiency (246).

An alternative to this process is low (<10^{–9} N m (10^{–2} dynes cm)) tension polymer flooding where lower concentrations of surfactant are used compared to micellar polymer flooding. Chemical adsorption is reduced compared to micellar polymer flooding. Increases in oil production compared to waterflooding have been observed in laboratory tests. The physical chemistry of this process has been reviewed (247). Among the surfactants used in this process are alcohol propoxyethoxy sulfonates, the structure of which can be adjusted to the salinity of the injection water (248).

Alkaline Flooding. Alkaline or caustic flooding involves injection of high pH agents such as sodium hydroxide, sodium carbonate, or sodium silicate solutions. At equivalent Na₂O levels, the three alkaline agents give equivalent recovery of each of nine different crude oils in laboratory core floods (249). However, the use of buffered sodium carbonate rather than strong alkali can result in reduced interaction with mineral surfaces. The lower reagent consumption can reduce the amount of sodium carbonate required.

These chemicals generate surfactants *in situ* by reacting with organic acids present in crude oil (250,251). Several oil recovery mechanisms may be operative. Probably the most significant oil recovery mechanism for this process is lowering of the capillary number (the ratio of viscous to capillary forces) through interfacial tension reduction. Other possible mechanisms are altering rock wettability (usually from oil-wet to water-wet), oil emulsification and entrapment resulting in lower injectant mobility (in turn resulting in a greater injectant volumetric sweep efficiency), oil emulsification and entrainment in the flowing aqueous phase, and possibly the solubilization of rigid films that may form at the oil–water interface.

Caustic flooding chemicals are relatively inexpensive. However, project economics are adversely affected by the large quantities that must be injected. The high pH agents react with reservoir clays (252) and are precipitated by divalent metal ions present in formation waters. Coinjection of a lignosulfonate (253) or a polyacrylate (254) reduces precipitation. This precipitation has been used to advantage to reduce adverse surfactant and polymer interactions with dissolved divalent metal ions. Injecting a caustic preflush causes divalent metal ion precipitation before beginning a micellar polymer flood (255). Ion-exchange processes promoting solubilization of divalent metal ions limit the effectiveness of preflushes injected before the caustic solution (256).

Including a surfactant in the caustic formulation (surfactant-enhanced alkaline flooding) can increase optimal salinity of a saline alkaline formulation. This can reduce interfacial tension and increase oil recovery (255,257,258). Encouraging field test results have been reported (259). Both nonionic and anionic surfactants have been evaluated in this application (260,261).

Surfactants evaluated in surfactant-enhanced alkaline flooding include internal olefin sulfonates (259,261), linear alkylxylene sulfonates (262), petroleum sulfonates (262), alcohol ethoxysulfates (258,261,263), and alcohol ethoxylates anionic surfactants (257). Water-thickening polymers, either xanthan or polyacrylamide, can reduce injected fluid mobility in alkaline flooding (264) and surfactant-enhanced alkaline flooding (259,263). The combined use of alkali, surfactant, and water-thickening polymer has been termed the alkali–surfactant–polymer (ASP) process. Cross-linked polymers have been used to increase volumetric sweep efficiency of surfactant–polymer–alkaline agent formulations (265).

Steam flooding (28,266,267) can greatly increase the recovery of high viscosity crude oils by heat thinning. This increases oil mobility in the reservoir. The addition of urea and iron sulfate or nickel compounds is said to further lower the viscosity of the crude oil (268). Surfactant foaming agents can be used to reduce the mobility of the high temperature steam. Because some heavy crude oils have relatively high acid numbers, it is not surprising that addition of alkaline agents to high temperature steam can increase recovery of these oils (269).

Other Technologies

Microbial-enhanced oil recovery involves injection of carefully chosen microbes. Subsequent injection of a nutrient is sometimes employed to promote bacterial growth. Molasses is the nutrient of choice owing to its low (ca \$100 t) cost. The main nutrient source for the microbes is often the crude oil in the reservoir. A rapidly growing microbe population can reduce the permeability of thief zones improving volumetric sweep efficiency. Microbes, particularly species of Clostridium and Bacillus, have also been used to produce surfactants, alcohols, solvents, and gases *in situ* (270). These chemicals improve waterflood oil displacement efficiency (see also BIOREMEDIATION (SUPPLEMENT)).

Microbes adsorb and grow on reservoir rock surfaces fed by injected nutrients (271) and may have application in plugging thief zones near injection

well bores. However, there is concern that the microbes can also enter lower permeability zones containing higher oil saturations thereby reducing oil production. Controlling the rate and location of bacterial growth and chemical production can be difficult. Bacterial growth near wellbores has been a common problem causing reduced injection rates and productivity. However, field test results have been promising (272).

The *in situ* combustion method of enhanced oil recovery through air injection (28,273,274) is a chemically complex process. There are three types of *in situ* combustion: dry, reverse, and wet. In the first, air injection results in ignition of crude oil and continued air injection moves the combustion front toward production wells. Temperatures can reach 300–650°C. Ahead of the combustion front is a 90–180°C steam zone, the temperature of which depends on pressure in the oil reservoir. Zones of hot water, hydrocarbon gases, and finally oil propagate ahead of the steam zone to the production well.

The oil zone is fairly cool, and in a viscous oil reservoir this can result in little oil movement (liquid blocking). Reverse combustion, in which oil ignition occurs near the production well, can avoid this problem. The combustion zone moves countercurrent to the flow of air from the injection well. Oil flows through heated rock and remains mobile. Reverse combustion requires more air and consumes more oil than forward combustion.

In wet combustion, water is injected concurrently and alternately with air, extending the steam zone and aiding heat transfer to the crude oil reducing oil viscosity. This can decrease injected air:produced oil ratio and improve project economics.

Maintenance and propagation of the combustion front are problems. This has led to a near-wellbore technology in which the same well is used for air injection and oil production. The combustion front needs to be propagated for a relatively short distance (275).

Economic Aspects

World EOR production at the beginning of 1994 was 3×10^5 m³ d (1.9×10^6 bbl d)³ (2). The annual value of this production as of 1994 was on the order of $\$10 \times 10^9$. High viscosity crude oil recovered by steam flooding has a lower value than the average crude. The annual value of U.S. EOR production in 1994 was ca $\$1.7 \times 10^9$. If oil prices remain at a value of $\$100\text{--}\125 m³ ($\$16\text{--}\20 bbl) EOR technology could recover an additional 1×10^9 m³ ($7\text{--}9 \times 10^9$ bbl) of oil in the United States (276).

There are a large number of EOR projects worldwide consuming large quantities of enhanced oil recovery chemicals. For example, the consumption of CO₂ in the west Texas Permian Basin EOR projects was about 31×10^6 m³ d in 1994 (2). China's Daqing Field is the largest polymer flood EOR project and a polymer plant having a production capacity of 65,000 t yr is being built to meet polymer injection requirements. This investment is predicted to increase oil production by 15,900 m³ d and add about 160×10^6 m³ (1×10^9 bbl) to the recoverable reserves from Daqing Field (2).

Whereas EOR processes using large volumes of materials combined with high concentrations of surfactants and polymers are often uneconomical, several processes using small volumes of chemicals to increase the volumetric sweep efficiency of waterflooding, supercritical CO₂ and gas injection EOR, and steam flooding are being used. Moreover, several cross-linked polymer systems have been shown to be effective in plugging fractures and substantially reducing the permeability of thief zones.

BIBLIOGRAPHY

"Enhanced Oil Recovery" under "Petroleum" in *ECT* 3rd ed., Vol. 17, pp. 168–182, by H. R. Froning, D. D. Fussell, and E. W. Heffern, Amoco Production Co.

1. M. H. Holtz, S. C. Ruppel, and C. R. Hocott, *J. Pet. Technol.* **44**, 1250 (1992).
2. G. Moritis, *Oil Gas J.*, 51 (Sept. 26, 1994).
3. *An Evaluation of Known Remaining Oil Resources in the United States*, Interstate Oil and Gas Compact Commission, Oklahoma City, Okla., Nov. 1993.
4. D. H. Smith, *Surfactant-Based Mobility Control—Progress in Miscible-Flood Enhanced Oil Recovery*, ACS Symposium Series No. 373, American Chemical Society, Washington, D.C., 1988.
5. F. F. Craig, Jr., *The Reservoir Engineering Aspects of Water Flooding*, Society of Petroleum Engineers, Monograph No. 3, Dallas, Tex., 1971, pp. 19–22.
6. **U.S. Pat. 5,320,170** (June 14, 1994), W. S. Huang, J. J. Hsu, and B. N. Wang (to Texaco, Inc.).
7. R. M. Butler, *J. Can. Pet. Tech.* **28**, 39 (Mar. 1989).
8. S. D. Joshi, *J. Pet. Technol.* **43**, 1302 (1991).
9. *Technical Constraints Limiting Application of Enhanced Oil Recovery Techniques to Petroleum Production in the United States*, U.S. Dept. of Energy, DOE BETC RI-83 9 (DEB4003910), Washington, D.C., Jan. 1984.
10. G. V. Chilinger and T. F. Yen, *Energy Sources* **7**, 67 (1983).
11. N. R. Morrow, *J. Pet. Tech.* **42**, 1476 (1990).
12. W. G. Anderson, *J. Pet. Technol.* **39**, 1453 (1987).
13. W. G. Anderson, *J. Pet. Technol.* **38**, 1125 (1986).
14. **U.S. Pat. 4,690,217** (Sept. 1, 1987), D. L. Taggart and E. W. Heffern (to Amoco Corp.).
15. P. F. Dymond and P. R. Spurr, *SPE Reservoir Eng.* **3**, 165 (1988).
16. D. M. Clementz, D. E. Patterson, R. J. Aseltine, and R. E. Young, *J. Pet. Technol.* **34**, 2087–2096 (1982).
17. F. Cusack, H. M. Lappin-Scott, and J. W. Costerton, *Oil Gas J.* **85**, 87 (Nov. 9, 1987).
18. W. N. Hensel, Jr., R. L. Sullivan, and R. H. Stallings, *Pet. Eng. Int.*, 155 (May 1981).
19. M. G. Reed, *J. Pet. Technol.* **24**, 860–864 (1972).
20. J. K. Borchardt, in J. K. Borchardt and T. F. Yen, eds., *Oilfield Chemistry: Enhanced Recovery and Production Stimulation*, American Chemical Society, Washington, D. C., 1989, p. 204.
21. R. D. Sydansk, *J. Pet. Technol.* **36**, 1366 (1984).
22. F. I. Stalkup, Jr., *Miscible Displacement*, Monograph Vol. 8, Society of Petroleum Engineers, Dallas, Tex., 1983, p. 62.
23. D. F. Martin and J. J. Tabor, *J. Pet. Technol.* **44**, 396 (1992).
24. L. Minssieux, in Ref. 20, p. 273.
25. R. T. Johansen, in *Improved Oil Recovery*, Interstate Oil Compact Commission, Oklahoma City, Okla., 1983, pp. 91–112.
26. C. Bardon and co-workers, *SPE Res. Eng.* **9**, 92 (1994).
27. F. H. Poettmann, in *Improved Oil Recovery*, Interstate Oil Compact Commission, Oklahoma City, Okla., 1983, p. 173.
28. M. Prats, *Thermal Recovery*, Monograph Vol. 7, Society of Petroleum Engineers, Dallas, Tex., 1982.
29. J. K. Borchardt, *Coll. Surf.* **63**, 189 (1992) and references therein.
30. R. D. Sydansk and P. E. Moore, *Oil Gas J.*, 40–45 (Jan. 20, 1992).
31. P. D. Moffitt, *J. Pet. Technol.* **45**, 356 (1993).
32. J. K. Borchardt, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 10, John Wiley & Sons, Inc., New York, 1987, pp. 350–351.
33. A. Moradi-Araghi and P. H. Doe, *SPE Reservoir Eng.* **2**, 189 (1987).
34. S. K. Nanda, R. Kumar, K. L. Sindhvani, and K. L. Goyl, *ONGC Bull.* **23**, 175 (1986).
35. **U.S. Pat. 5,244,936** (Sept. 14, 1993), T. O. Mitchell (to Mobil Oil Corp.).
36. Ref. 32, p. 344.
37. R. K. Prud'homme, J. T. Uhl, J. P. Poinsette, and F. Halverson, *Soc. Pet. Eng. J.* **23**, 804 (1983).
38. **U.S. Pat. 4,413,680** (Nov. 8, 1983), B. B. Sandiford, H. T. Dovan, and R. D. Hutchins (to Union Oil Co., Calif.).
39. B. Sloat, *Pet. Eng. Int.*, 20 (1977).

40. T. P. Lockhart and P. Albonico, *SPE Prod. Engr.* **9**, 273 (1994).

41. D. J. Jordan, D. W. Green, R. E. Terry, and G. P. Willhite, *Soc. Pet. Eng. J.* **22**, 463 (1982).

42. **U. S. Pat. 3,687,200** (Aug. 29, 1972), W. G. Routson (to Dow Chemical Co.).

43. N. A. Mumallah, *SPE Reservoir Eng.* **3**, 243 (1988).

44. **U. S. Pat. 4,744,418** (May 17, 1988), R. D. Sydansk (to Marathon Oil Co.).

45. J. C. Mack and J. Warren, *J. Pet. Technol.* **36**, 1145 (1984).

46. **U.S. Pat. 4,723,605** (Feb. 9, 1988), R. D. Sydansk (to Marathon Oil Co.).

47. C-G. Hyang, D. W. Green, and G. P. Willhite, *SPE Reservoir Engr.* **1**, 583 (1986).

48. **U.S. Pat. 4,155,405** (May 22, 1979), L. Vio (to Societe Nationale Elf Aquitaine).

49. **U.S. Pat. 4,125,478** (Nov. 14, 1978), E. J. Sullivan and G. D. Jones (to Dow Chemical Co.).

50. **U.S. Pat. 4,579,667** (Apr. 1, 1986), E. Ech and R. D. Lees (to Hercules, Inc.).

51. A. Moradi-Araghi, G. Bjornson, and P. H. Doe, *SPE Adv. Technol. Ser.* **1**(1), 140 (1993).

52. **Eur. Pat. Appl. 273,210** (Dec. 12, 1986), R. G. Ryles (to American Cyanamid Co.).

53. **U.S. Pat. 4,688,639** (Aug. 25, 1987), D. O. Falk (to Marathon Oil Co.).

54. S. Hejri, F. Jousset, D. W. Green, C. S. McCool, and G. P. Willhite, *SPE Res. Eng.* **8**, 229 (1993).

55. **U.S. Pat. 4,574,887** (Mar. 11, 1986), M. K. Abdo (to Mobil Oil Corp.).

56. **U.S. Pat. 4,439,334** (Mar. 27, 1984), J. K. Borchardt (to Halliburton Co.).

57. **U.S. Pat. 4,503,909** (Mar. 12, 1985), P. A. Argabright and J. S. Rhudy (to Marathon Oil Co.).

58. **U.S. Pat. 4,428,429** (Jan. 31, 1984), B. J. Felber and C. A. Christopher (to Standard Oil Co.).

59. **U.S. Pat. 3,334,689** (Oct. 8, 1967), H. C. McLaughlin (to Halliburton Co.).

60. **U.S. Pat. 4,321,968** (Mar. 30, 1982), E. E. Clear (to Phillips Petroleum Co.).

61. **U.S. Pat. 4,275,789** (June 30, 1981), D. D. Lawrence and B. J. Felber (to Standard Oil Co., Indiana).

62. **U.S. Pat. 4,299,690** (Nov. 11, 1980), B. W. Allan (Texaco, Canada, Inc.).

63. **Can. Pat. 1,187,404** (May 21, 1985), M. Navritil, J. P. Batycky, M. Sovak, and M. S. Mitchell (Borden Co., Ltd.).

64. P. M. Hess, *J. Pet. Technol.* **32**, 1834 (1980).

65. **U.S. Pat. 4,663,367** (May 5, 1987), M. Navritil, M. S. Mitchell, and M. Sovak (to Borden Co., Ltd.).

66. **Can. Pat. 1,217,932** (Feb. 17, 1987), M. Navritil, M. S. Mitchell, and M. Sovak (to Borden Co.).

67. **U.S. Pat. 4,745,976** (May 24, 1988), J. H. Harwell and J. F. Scamehorn (to Board of Regents for the University of Oklahoma).

68. **U.S. Pat. 4,161,983** (July 24, 1979), V. H. Schievelbein (to Texaco, Inc.).

69. R. D. Sydansk, *SPE Res. Eng.* **5**, 346 (1992).

70. **U.S. Pat. 4,460,627** (July 17, 1984), J. D. Weaver, L. E. Harris, and W. M. Harms (to Halliburton Co.).

71. A. Zaitoun, N. Kohler, and Y. Guerrini, *J. Pet. Technol.* **43**, 862 (1991).

72. K. S. Sorbie, *Polymer-Improved Oil Recovery*, Blackie and Son, Ltd., London, 1991.

73. Ref. 5, pp. 45–47.

74. B. C. Crafts and W. F. Hawkins, *Applied Petroleum Reservoir Engineering*, Prentice-Hall, New York, 1959.

75. J. K. Borchardt, in Ref. 32, pp. 328–369.

76. P. B. Lorenz, J. C. Trantham, P. R. Zornes, and C. G. Dodd, *SPE Res. Eng.* **1**, 341 (1986).

77. J. Chatterji and J. K. Borchardt, *J. Pet. Technol.* **32**, 2042–2056 (1981).

78. C. S. H. Chen and E. W. Sheppard, *J. Macromol. Sci. Chem. Part A*, 239 (1979).

79. D. M. Jones, K. Walters, and P. R. Williams, *Rheological Acta* **26**, 20 (1986).

80. **U.S. Pat. 4,053,699** (Oct. 11, 1977), P. T. Cahalan, J. A. Peterson, and D. A. Arndt (to General Mills Chemicals Inc.).

81. **U.S. 4,104,193** (Aug. 1, 1978), W. H. Carter, C. A. Christopher, and T. Jefferson (to Texaco Inc.).

82. P. A. Sandford and A. Laskin, eds., *Extracellular Microbial Polysaccharides*, ACS Symposium Series No. 45, American Chemical Society, Washington, D.C., 1977.

83. **U.S. Pat. 4,212,748** (July 15, 1980), B. M. Casad, D. Conley, H. H. Ferrell, and O. M. Stokke (to Conoco, Inc.).

84. **Brit. Pat. 2,115,430** (Sept. 7, 1983), J. R. Bragg (to Exxon Production Research Co.).

85. M. Rinaudo and M. Milas, *Int. J. Biol. Macromol.* **2**, 45 (1980).

86. N. Kohler, D. Longchamp, and M. Thery, *J. Pet. Technol.* **39**, 835 (1987).

87. **U.S. Pat. 4,165,257** (Aug. 21, 1979), O. M. Stokke (to Conoco, Inc.).

88. S. L. Wellington, *Soc. Pet. Eng. J.* **23**, 901 (1983).

89. M. C. Cadmus and co-workers, *Appl. Environ. Microbiol.*, 5 (Aug. 1982).

90. **Ger. Pat. 2,715,026** (Oct. 13, 1977), S. L. Lee (to Shell Internationale Research Maatschappij BV).

91. **Eur. Pat. 106,666** (Apr. 24, 1984), C. J. Philips (to Pfizer Corp.).

92. **Can. Pat. 1,070,492** (Jan. 29, 1980), S. L. Wellington (to Shell Canada, Ltd.).

93. **U.S. Pat. 4,454,620** (June 12, 1984), F. Dawans, D. Binet, N. Kohler, and D. V. Quang (to Institut Francais du Petrole).

94. **U.S. Pat. 4,481,316** (Nov. 6, 1984), S. Kanda and G. Kawamura (to Nitto Chemical Industry Co.).

95. **Jpn. Kokai Tokkyo Koho 61 275,337** (Dec. 5, 1986), S. Kanda and G. Kawamura (to Nitto Chemical Industry Co.).

96. **Jpn Kokai Tokkyo Koho 62 184,048** (Aug. 12, 1987), A. Niita, T. Arai, R. Funato, and T. Sato (to Mitsui Cyanamid KK).

97. **Jpn Kokai Tokkyo Koho 62 184,047** (Aug. 12, 1987), A. Niita, T. Arai, R. Funato, and T. Sato (to Mitsui Cyanamid KK).

98. **Jpn Kokai Tokkyo Koho 62 177,052** (Aug. 3, 1987), A. Niita, T. Arai, R. Funato, and T. Sato (to Mitsui Cyanamid KK).

99. **Jpn Kokai Tokkyo Koho 62 277,407** (Dec. 2, 1987), Y. Nakanishi (Mitsubishi Chemical Industries KK).

100. **Eur. Pat. Appl. 196,199** (Oct. 1, 1986), C. P. Judson (to Pfizer, Inc.).

101. **Jpn Kokai Tokkyo Koho 61 136,545** (June 24, 1986), A. Niita and Y. Ito (to Mitsui Toatsu Chemicals, Inc.).

102. **Eur. Pat. Appl. 241,340** (Oct. 14, 1987), F. Contat and J. Boutin (to Rh ne Poulenc Chemicals).

103. **Ger. Pat. 2,848,984** (May 17, 1979), W. C. Wernau (to Pfizer Corp.).

104. **Eur. Pat. 103,483** (Mar. 21, 1984), G. M. Holzwarth (to Exxon Research and Engineering Co.).

105. J. C. Phillips, J. W. Miller, W. C. Wernau, B. E. Tate, and M. H. Auerbach, *Soc. Pet. Eng. J.* **25**, 594 (1985).

106. **U.S. Pat. 4,457,372** (July 3, 1984), M. S. Doster, A. J. Nute, and C. A. Christopher (to Texaco, Inc.).

107. **Eur. Pat. Appl. 211,288** (Feb. 25, 1987), R. W. Vanderslice and P. Shanon (to Getty Science and Development Co.).

108. A. Audibert, C. Noik, and J. Lecourtier, *J. Can. Pet. Technol.* **32**(7), 53 (1993).

109. **U.S. Pat. 4,439,332** (Mar. 27, 1984), S. Frank, A. T. Coscia, and J. M. Schmitt (to American Cyanamid Co.).

110. **U.S. Pat. 4,439,342** (May 8, 1984), J. K. Borchardt (to Halliburton Co.).

111. J. Klein and R. Heitzman, *Makromol. Chem.* **179**, 1895 (1978).

112. F. H. Halverson and J. E. Lancaster, *Macromolecules* **18**, 1139 (1985).

113. **Fr. Pat. 2,495,217** (June 4, 1982), J. Boutin and F. Contat (to Rh ne Poulenc Industries).

114. G. Muller, J. P. Laine, and J. C. Fenyo, *J. Polym. Sci. Polym. Chem. Ed.* **17**, 659 (1979).

115. J. S. Ward and F. D. Martin, *Soc. Pet. Eng. J.* **21**, 623 (1981).

116. R. G. Ryles, *SPE Reservoir Eng.* **3**, 23 (1988).

117. A. Moradi-Araghi and P. H. Doe, *SPE Reservoir Eng.* **2**, 189 (1987).

118. M. Szabo, *J. Pet. Technol.* **31**, 553 (1979).

119. M. Szabo, *J. Pet. Technol.* **31**, 561 (1979).

120. J. G. Southwick and C. W. Manke, *SPE Reservoir Eng.* **3**, 1193 (1988).

121. C. L. McCormick, *J. Macromol. Sci. Chem. Part A.* **22**, 955 (1985).

122. **U.S. Pat. 4,644,020** (Feb. 17, 1987), G. A. Stahl (to Phillips Petroleum Co.).

123. P. H. Doe, A. Moradi-Araghi, J. E. Shaw, and G. A. Stahl, *SPE Res. Eng.* **2**, 461 (1987).

124. C. L. McCormick and K. P. Blackmon, *J. Polym. Sci. A, Polym. Chem.* **24**, 2635 (1986).

125. **Ger. Pat. 3,627,456** (Feb. 26, 1987), T. Y. Ching (to Chevron Research Co.).

126. **Eur. Pat. Appl. 233,533** (Aug. 26, 1987), R. G. Ryles (to American Cyanamid Co.).

127. **Ger. Pat. 3,220,503** (Dec. 1, 1983), P. Engelhardt, U. Greiner, H. Schmitz, W. Gulden, and S. P. von Halasz (to Casella AG).

128. **Jpn Kokai Tokkyo Koho 62 015,279** (Jan. 23, 1987), Y. Ito, A. Niita, Y. Sudo, and K. Hayashi (to Mitsui Toatsu Chemicals, Inc.).

129. **U.S. Pat. 4,540,498** (Sept. 10, 1985), M. M. Wu and L. E. Ball (Standard Oil Co., Ohio).

130. R. W. Dexter and R. G. Ryles, in Ref. 20, p. 102.

131. H. Pledger, Jr., J. J. Meister, T. E. Hogen-Esch, and G. B. Butler, *Polym. Prepr. Am. Chem. Soc. Div. Polym. Chem.* **22**(2), 72 (1981).

132. H. H. Neidlinger and C. L. McCormick, *Polym. Prepr. Am. Chem. Soc. Div. Polym. Chem.* **20**(1), 901 (1979).

133. J. J. Meister, *Polym. Prepr. Am. Chem. Soc. Div. Polym. Chem.* **25**(1), 266 (1984).

134. **U.S. Pat. 4,524,003** (June 18, 1985), J. K. Borchardt (to Halliburton Co.).

135. **U.S. Pat. 3,919,092** (Nov. 11, 1975), C. J. Norton and D. D. Falk (to Marathon Oil Corp.).

136. **U.S. Pat. 4,508,629** (Apr. 2, 1985), J. K. Borchardt (to Halliburton Co.).

137. **U.S. Pat. 4,039,028** (Aug. 7, 1977), R. K. Knight (to Union Oil Co., Calif.).

138. P. Kovacs, *Food Technol.* **27**, 26 (1973).

139. D. N. Schulz, D. A. Kaladas, J. J. Maurer, J. Bock, S. J. Pace, and W. W. Schulz, *Polymer* **28**, 2110 (1987).

140. **U.S. Pat. 4,694,046** (Sept. 15, 1987), J. Bock, D. B. Siano, and R. Turner (to Exxon Research and Engineering Co.).

141. **U.S. Pat. 4,730,028** (Mar. 8, 1988), J. Bock and P. L. Valint (to Exxon Research and Engineering Co.).

142. **U.S. Pat. 4,709,759** (Dec. 1, 1987), J. Bock, S. J. Pace, and D. N. Schulz (to Exxon Research and Engineering Co.).

143. S. Evani and G. D. Rose, *Polym. Mat. Sci. Eng.* **57**, 477 (1987).

144. **U.S. Pat. 4,409,110** (Oct. 11, 1983), J. K. Borchardt and D. L. Brown (to Halliburton Co.).

145. P. J. Shuler, D. L. Kuehne, J. T. Uhl, and G. W. Walkup, Jr., *SPE Reservoir Eng.* **2**, 271 (1987).

146. F. Friedmann, *SPE Reservoir Eng.* **1**, 261 (1986).

147. Y. Cohen and F. R. Christ, *SPE Reservoir Eng.* **1**, 113 (1986).

148. L. T. Lee, J. Lecourtier, and G. Chaveteau, in Ref. 20, p. 224.

149. S. A. Hong and J. H. Bae, *SPE Res. Eng.* **5**, 467 (1990).

150. **U.S. Pat. 4,627,494** (Dec. 9, 1986), G. Kalfoglou (to Texaco, Inc.).

151. **U.S. Pat. 4,713,185** (Dec. 15, 1987), J. Howard and M. Stirling (to Canadian Patents and Development Ltd.).

152. **U.S. Pat. 4,284,517** (Aug. 18, 1981), C. S. H. Chen and E. W. Sheppard (to Mobil Oil Corp.).

153. **Brit. Pat. 2,148,356** (Mar. 4, 1987), J. H. Clint, P. K. G. Hodgson, and E. J. Tinley (to British Petroleum PLC).

154. **U.S. Pat. 4,574,885** (Mar. 11, 1986), R. L. Horton (to Phillips Petroleum Co.).

155. T. G. Arf and co-workers, *SPE Reservoir Eng.* **2**, 166 (1987).

156. S. L. Wellington and H. J. Vinegar, *J. Pet. Technol.* **39**, 885 (1987).

157. G. J. Parlani, K. A. McColloch, S. L. Warden, and D. R. Edens, *J. Pet. Technol.* **44**, 984 (1992).

158. F. I. Stalkup, Jr., *J. Pet. Technol.* **36**, 815–826 (1984).

159. J. P. Heller, D. K. Dandge, R. J. Card, and L. G. Donaruma, *Soc. Pet. Eng. J.* **25**, 679 (1985).

160. J. P. Heller, in L. L. Schramm, ed., *Foams: Fundamentals and Applications in the Oil Industry*, American Chemical Society, Washington, D.C., 1994, p. 201.

161. K. Mannhardt and J. J. Novosad, in Ref. 160, p. 259.

162. J. E. Hanssen and T. Holt, *Rogoland Research Report*, RF-80 93, Stavanger, Norway, 1993, and references therein.

163. **U.S. Pat. 3,185,634** (May 25, 1965), F. F. Craig, Jr. and J. L. Lummus (to Pan American Petroleum Corp.).

164. M. I. Kuhlman, A. H. Falls, S. K. Hara, T. G. Monger-McClure, and J. K. Borchardt, *SPE Res. Eng.* **7**, 445 (1992).

165. J. K. Borchardt, in Ref. 20, pp. 181–204.

166. **U.S. Pat. 4,706,752** (Nov. 17, 1987), L. W. Holm (to Union Oil Co. of California).

167. **U.S. Pat. 4,706,750** (Nov. 17, 1987), J. J. Buckles (to Mobil Oil Corp.).

168. **U.S. Pat. 4,502,538** (Mar. 5, 1985), S. L. Wellington, J. Reisberg, E. F. Lutz, and D. B. Bright (to Shell Oil Co.).

169. **U.S. Pat. 4,799,547** (Jan. 24, 1989), J. K. Borchardt (to Shell Oil Co.).

170. **U.S. Pat. 4,380,266** (Apr. 19, 1983), S. L. Wellington (to Shell Oil Co.).

171. **U.S. Pat. 4,739,831** (Apr. 26, 1988), L. A. Settlemyer and M. J. McCoy (to Dow Chemical Co.).

172. M. K. Sharma, D. O. Shah, and W. E. Brigham, *SPE Reservoir Eng.* **1**, 253 (1986).

173. **U.S. Pat. 5,325,920** (July 5, 1994), N. F. Djabbarah (to Mobil Oil Corp.).

174. **U.S. Pat. 4,676,316** (June 30, 1987), T. O. Mitchell (to Mobil Oil Corp.).

175. **Can. Pat. 1,221,305** (May 5, 1987), N. F. Djabbarah (to Cities Service Oil and Gas Corp.).

176. **U.S. Pat. 4,733,727** (Mar. 29, 1988), A. H. Falls (to Shell Oil Co.).

177. L. W. Holm, *J. Pet. Technol.* **32**, 1499 (1970).

178. L. W. Holm and W. H. Garrison, *SPE Reservoir Eng.* **3**, 112 (1988).

179. L. D. Talley, *SPE Reservoir Eng.* **3**, 235 (1988).

180. **Brit. Pat. 2,239,278** (June 21, 1991), L. L. Schramm, C. Ayasse, K. Mannhardt, and J. Novosad (to Alberta Oil Sands Technology and Research Authority).

181. G. P. Small, *SPE Prod. Eng.* **1**, 388 (1986).

182. K. C. Hong and C. E. Stevens, *SPE Reservoir Eng.* **7**, 407 (1992).

183. T. J. C. Fontaine, L. Hayes, and G. Reese, *J. Can. Pet. Technol.* **32**, 44 (Nov. 1993).

184. P. J. Jespersen and T. J. C. Fontaine, *J. Can. Pet. Technol.* **32**, 52 (May 1993).

185. D. K. Olsen and P. S. Sarathi, *Proceedings of the Field Application of Foams for Oil Production Symposium*, U.S. Dept. of Energy, N.I.P.E.R., Bartlesville, Okla., 1993.

186. **U.S. Pat. 5,333,687** (Aug. 2, 1994), W. T. Osterdoh (to Texaco, Inc.).

187. E. E. Isaacs, F. C. McCarthy, and J. D. Maunder, *SPE Reservoir Eng.* **3**, 565 (1988).

188. **U.S. Pat. 4,693,311** (Sept. 15, 1987), H. M. Muijs and P. P. M. Keijzer (to Shell Oil Co.).

189. **Ger. Pat. 3,510,765** (Sept. 26, 1985), R. Janssen van Rosmalen, H. M. Muijs, and P. P. M. Keijzer (to Shell Internationale Research Maatschappij BV).

190. **U.S. Pat. 5,273,682** (Dec. 28, 1993), M. Danzik (to Chevron Research and Technology Co.).

191. **Ger. Pat. 3,734,075** (Apr. 4, 1988), H. M. Muijs and P. P. M. Keijzer (to Shell Internationale Research Maatschappij BV).

192. **Ger. Pat. 3,503,532** (Aug. 8, 1985), W. S. Huang, Z. Z. Gassmann, J. T. Hawkins, V. H. Schievelbein, and W. L. Hall (to Texaco Development Co.).

193. **U.S. Pat. 4,699,214** (Oct. 13, 1987), H. P. Angstadt (to Sun Refining and Marketing Co.).

194. **PCT Int. Pat. Appl. 85 05,146** (1985), T. Lim (to Dow Chemical Co.).

195. **U.S. Pat. 4,576,232** (Mar. 18, 1986), J. H. Duerksen, R. G. Wall, and J. D. Knight (to Chevron Research Co.).

196. A. H. Falls, J. B. Lawson, and G. J. Hirasaki, *J. Pet. Technol.* **40**, 95 (1988).

197. **U.S. Pat. 4,570,711** (Feb. 18, 1986), A. H. Falls, J. B. Lawson, and G. J. Hirasaki (to Shell Oil Co.).

198. J. Novosad, B. B. Maini, and A. Huang, *J. Can. Pet. Technol.* **25**, 42 (1986).

199. A. Y. Huang and J. Novosad, *Proc. Eng. Found. Conf. Fundam. Adsorpt., 2nd, 1986*, Engineering Foundation, New York, 1987, pp. 265–275.

200. H. C. Lau and S. M. O'Brien, *SPE Reservoir Eng.* **2**, 1177 (1988).

201. **U.S. Pat. 4,597,442** (July 1, 1986), R. E. Dilgren, H. C. Lau, and G. J. Hirasaki (to Shell Oil Co.).

202. **U.S. Pat. 4,617,995** (Oct. 21, 1986), H. C. Lau (to Shell Oil Co.).

203. **Fr. Pat. 2,557,198** (Juen 28, 1985), H. P. Angstadt, D. F. Rugen, and J. L. Cayias (to Sun Marketing and Refining Co.).

204. E. E. Isaacs, J. D. Maunder, and L. Juan, in Ref. 20, p. 325.

205. B. B. Maini and V. Ma, *J. Can. Pet. Technol.* **25**, 65–69 (1986).

206. H. P. Angstadt and H. Tsao, *SPE Reservoir Eng.* **2**, 613 (1987).

207. **U.S. Pat. 4,702,317** (Oct. 27, 1987), C. W. Shen (to Texaco, Inc.).

208. **U.S. Pat. 4,727,938** (Mar. 1, 1988), H. C. Lau (to Shell Oil Co.).

209. H. C. Lau and J. K. Borchardt, *SPE Reservoir Eng.* **6**, 470 (1991).

210. **U.S. Pat. 5,105,884** (Apr. 21, 1992), R. D. Sydansk (to Marathon Oil Co.).

211. D. O. Shah and R. S. Schechter, eds. *Improved Oil Recovery by Surfactant and Polymer Flooding*, Academic Press, Inc., New York, 1977.

212. H. K. van Poolen, *Fundamentals of Enhanced Oil Recovery*, PennWell Publishing Co., Tulsa, Okla., 1980, p. 114.

213. J. C. Morgan, R. S. Schechter, and W. H. Wade, in K. L. Mittal, ed., *Proc. Sect. 52nd Colloid Surf. Sci. Symp.*, Vol. 2, Plenum Press, New York, 1979, pp. 749–775.

214. M. Bourrel, F. Verzaro, and C. Chambu, *SPE Res. Eng.* **2**, 41 (1987).

215. A. Graciaa, L. M. Fortney, R. S. Schechter, W. H. Wade, and S. Yiv, *Soc. Pet. Eng. J.* **20**, 743 (1982).

216. **Ger. Pat. 3,542,063** (June 12, 1986), M. C. Puerto (to Exxon Production and Research Co.).

217. **U.S. Pat. 4,270,607** (June 2, 1981), R. L. Cardenas, B. G. Harnsberger, and J. Maddox, Jr. (to Texaco, Inc.).

218. Y. Barakat and co-workers, *Soc. Pet. Eng. J.* **23**, 913 (1983).

219. **U.S. Pat. 4,555,351** (Nov. 26, 1985), H. Morita, Y. Kawada, J. Yamada, and T. Ukigai (to Lion Corp.).

220. **U.S. Pat. 4,512,404** (Apr. 23, 1985), H. Morita, Y. Kawada, J. Yamada, and T. Ukigai (to Lion Corp.).

221. **Brit. Pat. 2,160,242** (Sept. 9, 1987), R. Thaver (to British Petroleum PLC).

222. **Brit. Pat. Appl. 2,184,763** (July 1, 1987), R. Thaver (to British Petroleum PLC).

223. **Ger. Pat. 3,303,894** (Aug. 8, 1983), H. S. Aldrich, T. L. Ashcraft, M. C. Puerto, and R. L. Reed (to Exxon Production Research Co.).

224. **Ger. Pat. 3,422,613** (Dec. 19, 1985), N. Greif, K. Oppenlaender, and K. U. Sewe (to BASF AG).

225. **Eur. Pat. Appl. 264,867** (Apr. 27, 1988), R. Schmidt, W. Rupp, G. Schneider, and E. M. Kohn (to Hoechst AG).

226. **U.S. Pat. 4,426,302** (Jan. 17, 1984), D. R. McCoy, R. M. Gipson, and C. G. Naylor (to Texaco, Inc.).

227. Y. C. Chiu and H. J. Hwang, *Colloids Surf.* **28**, 53 (1987).

228. J. E. Shaw, *J. Am. Oil Chem. Soc.* **61**, 1395 (1984).

229. **U.S. Pat. 4,692,551** (Sept. 8, 1987), K. C. Liu (to GAF Corp.).

230. M. Abe, R. S. Schechter, R. D. Selliah, B. Sheikh, and W. H. Wade, *J. Dispersion Sci. Technol.* **8**, 157 (1987).

231. M. Baviere, B. Bazin, and C. Noil, *SPE Reservoir Eng.* **3**, 597 (1988).

232. M. C. Puerto and R. L. Reed, *Soc. Pet. Eng. J.* **23**, 669 (1983).

233. **Fr. Pat. 2,589,858** (May 15, 1987), H. S. Aldrich, M. C. Puerto, and R. L. Reed (to Exxon Production Research Co.).

234. C. Lalanne-Cassou and co-workers, *J. Dispersion Sci. Technol.* **8**, 137 (1987).

235. T. A. B. M. Bolsman and G. J. R. Daane, *SPE Reservoir Eng.* **1**, 53 (1986).

236. B. B. Bansal, V. Hornof, and G. Neale, *Can. J. Chem. Eng.* **57**, 203 (1979).

237. K. Manasrah, G. H. Neale, and V. Hornof, *Cellul. Chem. Technol.* **19**, 291 (1985).

238. **U.S. Pat. 4,739,040** (Apr. 19, 1988), D. G. Naae, L. E. Whittington, W. A. Ledoux, and F. E. Debons (to Texaco, Inc.).

239. **U.S. Pat. 4,739,041** (Apr. 19, 1988), L. R. Morrow, M. G. Dague, and L. E. Whittington (to Texaco, Inc.).

240. D. B. Babu, G. Neale, and V. Hornof, *Cellul. Chem. Technol.* **20**, 663 (1986).

241. O. S. Hjelmeland and L. E. Larrondo, *SPE Reservoir Eng.* **1**, 321 (1986).

242. P. Somasundaran and H. H. Shafick, *Soc. Pet. Eng. J.* **25**, 343 (1985) and references therein.

243. P. B. Lorenz, J. C. Trantham, D. R. Zornes, and C. G. Dodd, *SPE Reservoir Eng.* **1**, 341 (1986).

244. J. S. Hohnson and R. R. M. Jones, *Ann. Chim. (Rome)* **77**, 229 (1987).

245. J. Ruboi and J. A. Kitchener, *J. Colloid Interface Sci.* **57**, 132 (1976).

246. P. H. Krumine, I. B. Ailin-Pyzik, and J. S. Falcone, Jr., *Prepr. Am. Chem. Soc., Div. Pet. Chem.* **26**(1), 195 (1981).

247. C. Methemitis, M. Morcellet, J. Sabbadin, and J. Francois, *Eur. Polym. J.* **22**, 619 (1986) and references therein.

248. T. Austad, I. Fjelde, K. Veggeland, and K. Taugboel, *J. Pet. Sci. Eng.* **10**(3), 255–269 (1994).

249. **U.S. 5,311,943** (May 17, 1994), S. L. Wellington (to Shell Oil Co.).

250. J. K. Burk, *SPE Reservoir Eng.* **2**, 9 (1987).

251. Ref. 213, pp. 104–111.

252. J. D. Neil, H. L. Chang, and T. M. Geffern, in *Improved Oil Recovery*, Interstate Oil Compact Commission, Oklahoma City, Okla., 1983, pp. 52–66.

253. S. M. Mohnot, J. H. Bae, and W. L. Foley, *SPE Reservoir Eng.* **2**, 653 (1987).

254. **U.S. Pat. 4,714,113** (Dec. 12, 1987), S. M. Mohnot and P. M. Chakrabarti (to PPG Industries, Inc.).

255. L. W. Holm and S. D. Robertson, *J. Pet. Technol.* **33**, 161 (1981).

256. **U.S. Pat. 4,466,892** (Aug. 21, 1984), K. S. Chan and S. J. Majoros (to Standard Oil Co., Ohio).

257. A. L. Bunge and C. J. Radke, *Soc. Pet. Eng. J.* **23**, 657 (1983).

258. P. J. Shuler, D. L. Kuehne, and R. M. Lerner, *J. Pet. Technol.* **41**, 80 (1989) and references therein.

259. R. C. Nelson and co-workers, *SPE Res. Eng.* **9**, 217 (1994).

260. R. C. Nelson, J. B. Lawson, D. R. Thigpen, and G. L. Stegemeier, *Proc. 1984 SPE Enhanced Oil Recovery Symposium*, Society of Petroleum Engineers, Dallas, Tex., 1984, pp. 417–424.

261. S. M. Saleem and M. J. Faber, *Rev. Tec. INTEVEP* **6**, 133–142 (1986).

262. S. M. Saleem and A. Hernandez, *J. Surf. Sci. Technol.* **3**, 1 (1987).

263. M. J. Pitts, *Prepr. Am. Chem. Soc., Div. Pet. Chem.*, **33** (1), 169 (1988).

264. **U.S. Pat. 4,502,541** (Mar. 5, 1985), J. B. Lawson and D. R. Thigpen (to Shell Oil Co.).

265. G. H. Neale, K. C. Khulbe, and V. Hornof, *Can. J. Chem. Eng.* **65**, 700 (1987).

266.

Ref. 213, pp. 3–40.

267.

S. M. Farouq Ali and R. F. Meldau, in *Improved Oil Recovery*, Interstate Oil Compact Commission, Oklahoma City, Okla., 1983, pp. 311–350.

268.

U.S. Pat. 5,314,615 (May 24, 1994), R. E. Campos and J. A. Hernandez (to Intevep, SA).

269.

D. M. Babu, V. Hornof, and G. Neale, *Can. J. Chem. Eng.* **62**, 156 (1984).

270.

G. E. Jenneman, R. M. Knapp, M. J. McInerney, D. E. Menzie, and D. E. Revus, *Soc. Pet. Eng. J.* **24**, 31 (1984).

271.

D. O. Hitzman, *Microbial Enhancement of Oil Recovery—Recent Advances*, Elsevier Science Publishing Co., Inc., New York, 1991, pp. 11–20.

272.

I. Lazar, *Dev. Petr. Sci.* **31**, 485 (1991).

273.

P. B. Crawford and C. Cju, *Improved Oil Recovery*, Interstate Oil Compact Commission, Oklahoma City, Okla., 1983, pp. 251–309.

274.

Ref. 213, pp. 41–55.

275.

B. I. Nzekwu, R. J. Hallam, and G. J. J. Williams, *SPE Reservoir Eng.* **5**, 163 (1990); Ref. 226, pp. 41–55.

276.

An Evaluation of Known Remaining Oil Reserves in the United States, Interstate Oil Compact Commission, Oklahoma City, Okla., Nov. 1993.

John K. Borchardt
Shell Development Company

REFINERY PROCESSES, SURVEY

Petroleum refining, also called petroleum processing, is the recovery and or generation of usable or salable fractions and products from crude oil, either by distillation or by chemical reaction of the crude oil constituents under the effects of heat and pressure. Synthetic crude oil, produced from tar sand (oil sand) bitumen, and heavier oils are also used as feedstocks in some refineries. Heavy oil conversion (1), as practiced in many refineries, does not fall into the category of synthetic fuels (syncrude) production. In terms of liquid fuels from coal and other carbonaceous feedstocks, such as oil shale (qv), the concept of a synthetic fuels industry has diminished over the past several years as being uneconomical in light of current petroleum prices.

Crude petroleum is a mixture of compounds boiling at different temperatures that can be separated into a variety of different generic but often overlapping fractions (Table 1). The amounts of these fractions produced by distillation depend on the origin and properties of crude petroleum (2).

Table 1. Distillation Fractions of Petroleum

Fraction	Boiling, °C
light naphtha	1–150
gasoline	1– 180
heavy naphtha	150–205
kerosene	205–260
stove oil	205–290
light gas oil	260–315
heavy gas oil	315–425
lubricating oil	>400
vacuum gas oil	425–600
residuum	>600

When petroleum occurs in a reservoir that allows the crude material to be recovered by pumping operations as a free-flowing dark-to-light colored liquid, it is often referred to as conventional petroleum. In some oil fields, the downhole pressure is sufficient for recovery without the need for pumping. Heavy oil differs from conventional petroleum in that its flow properties are reduced and it is much more difficult to recover from the subsurface reservoir. These materials have a much higher viscosity and lower API (American Petroleum Institute) gravity than conventional petroleum, and primary recovery of these petroleum types usually requires thermal stimulation of the reservoir.

Heavy oil generally has an API gravity of less than 20 degrees and usually, but not always, a sulfur content of >2% by weight. Extra heavy oil occurs in the near-solid state and is virtually incapable of free flow under ambient conditions. Bitumen, often referred to as native asphalt, is a subclass of extra heavy oil and is frequently found as the organic filling in pores and crevices of sandstones, limestones, or argillaceous sediments, in which case the organic and associated mineral matrix is known as rock asphalt.

A residuum, often shortened to resid, is the residue obtained from petroleum after nondestructive distillation has removed all the volatile materials. The temperature of the distillation is usually below 345°C because the rate of thermal decomposition of petroleum constituents is substantial above 350°C. Temperatures as high as 425°C can be employed in vacuum distillation. When such temperatures are employed and thermal decomposition occurs, the residuum is usually referred to as pitch. By inference, the name is used in the same manner as when it refers to the nonvolatile residue from the thermal decomposition of coal tar (3).

Asphalt, prepared from petroleum, often resembles native asphalt. When asphalt is produced by distillation, the product is called residual, or straight-run, asphalt. However, if the asphalt is prepared by solvent extraction of residua or by light hydrocarbon (propane) precipitation, or if it is blown or otherwise treated, the name should be modified accordingly to qualify the product, eg, propane asphalt.

Sour and sweet are terms referring to a crude oil's approximate sulfur content, which relates to odor. A crude oil that has a high sulfur content usually contains hydrogen sulfide, H₂S, and or mercaptans, RSH; it is called sour. Without this disagreeable odor, the crude oil is judged sweet.

History

The use of petroleum or derived materials, such as asphalt, and the heavier nonvolatile crude oils is an old art (2). In fact, petroleum utilization has been documented for more than five thousand years. The earliest documented uses occurred in Mesopotamia (ancient Iraq) when it was recognized that the nonvolatile derivatives (bitumen or natural asphalt and manufactured asphalt) could be used for caulking and as an adhesive for jewelry or as a mastic for construction purposes. There is also documented use of bitumen for medicinal use.

Approximately two thousand years ago, Arabian scientists developed methods for the distillation of petroleum, which were introduced into Europe by way of the Arabian incursions into Spain. Petroleum, used in China since it was encountered when drilling for salt, appears in documents of the third century. The Baku region of northern Persia was also reported by Marco Polo in 1271–1273 as having a commercial petroleum industry.

Interest in naphtha (nafta) began with the discovery that petroleum could be used as an illuminant and as a supplement to bituminous incendiaries, which were becoming increasingly common in warfare. Greek fire was a naphtha–bitumen (or naphtha–asphalt) mix; the naphtha provided the flame and the bitumen (or asphalt) provided the adhesive properties that prolonged the incendiary effect.

Modern refining began in 1859 with the discovery of petroleum in Pennsylvania. After completion of the first well, the surrounding areas were immediately leased and extensive drilling took place. In the post-1945 era, Middle Eastern countries continued to rise in importance because of new discoveries of vast reserves. The United States, though continuing to be the biggest producer, was also the principal consumer and thus was not an exporter of oil. At this time, oil companies began to roam much farther in the search for oil, which has resulted in significant discoveries in Europe, Africa, and Canada.

The impetus to develop the petroleum refining industry came from several changes in life-styles. The increased needs for illuminants, for fuel to drive the factories of the industrial revolution, for gasoline to power the automobiles, as well as the demand for aviation fuel, all contributed to the increased use of petroleum.

The product slate has also changed. The increased demand for gasoline and lubricants brought about an emphasis on refining crude oil. This, in turn, brought about changes in the way crude oil was refined and led to innovations and developments in the refining industry, thereby giving birth to the integrated petroleum refinery (Fig. 1).

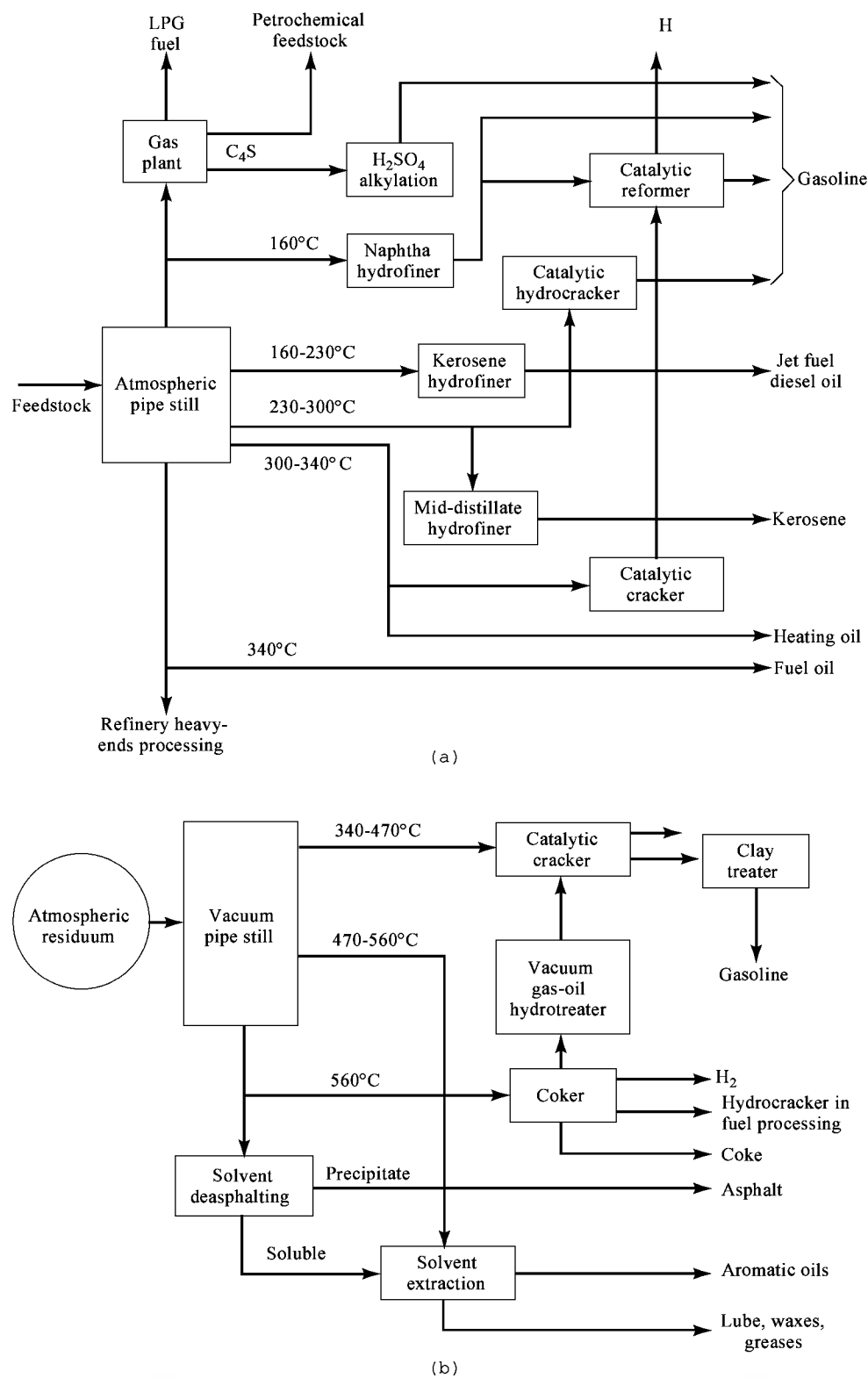


Fig. 1. General refinery operations: (a) light petroleum refining section; (b) heavy feedstock refining section.

Desalting and Dewatering

Crude oil is recovered from the reservoir mixed with a variety of substances: gases, water, and dirt (minerals) (4). Thus, refining actually commences with the production of fluids from the well or reservoir and is followed by pretreatment operations that are applied to the crude oil either at the refinery or prior to transportation. Pipeline operators, for instance, are insistent upon the quality of the fluids put into the pipelines; therefore, any crude oil to be shipped by pipeline or, for that matter, by any other form of transportation must meet rigid specifications in regard to water and salt content. In some instances, sulfur content, nitrogen content, and viscosity may also be specified.

Field separation, which occurs at a field site near the recovery operation, is the first attempt to remove the gases, water, and dirt that accompany crude oil coming from the ground. The separator may be no more than a large vessel that gives a quieting zone for gravity separation into three phases: gases, crude oil, and water containing entrained dirt.

Desalting is a water-washing operation performed at the production field and at the refinery site for additional crude oil cleanup. If the petroleum from the separators contains water and dirt, water washing can remove much of the water-soluble minerals and entrained solids. If these crude oil contaminants are not removed, they can cause operating problems during refinery processing, such as equipment plugging and corrosion as well as catalyst deactivation.

The usual practice is to blend crude oils of similar characteristics, although fluctuations in the properties of the individual crude oils may cause significant variations in the properties of the blend over a period of time. Blending several crude oils prior to refining can eliminate the frequent need to change the processing conditions that may be required to process each of the crude oils individually.

However, simplification of the refining procedure is not always the end result. Incompatibility of different crude oils, which can occur if, for example, a paraffinic crude oil is blended with a heavy asphaltic oil, can cause sediment formation in the unrefined feedstock or in the products, thereby complicating the refinery process (5).

Feedstock Evaluation

Three frequently specified properties are density–specific gravity–API gravity, characterization factor, and sulfur content (2,6,7). The API (American Petroleum Institute) gravity is a measure of density or specific gravity (sp gr):

$$^{\circ}\text{API} = \frac{141.5}{\text{sp gr}} - 131.5$$

Specific gravity is the ratio of the weight of a given volume of oil to the weight of the same volume of water at a standard temperature, usually 60°F (15.6°C). This method of measuring density and gravity first arose as a result of the need to define the character of products in more detail; it was natural to extend the measure to crude oils in general.

The Watson characterization factor has also been used as a measure of the chemical character of a crude oil or its fractions:

$$K = (T_B)^{1/3} \text{ sp gr}$$

where T_B is the absolute boiling point in degrees Rankine ($^{\circ}\text{R} = \frac{9}{5}\text{K}$) and sp gr is specific gravity compared to water at 60°F (15.6°C).

For a wide-boiling-range material such as crude oil, the boiling point is taken as an average of the five temperatures at which 10, 30, 50, 70, and 90% of the material is vaporized. A highly paraffinic crude oil can have a characterization factor as high as 13, whereas a highly naphthenic crude oil can be as low as 10.5, and the breakpoint between the two types of crude oil is approximately 12.

Refining

A refinery is a group of manufacturing plants that vary in number according to the variety of products produced (2). Refinery processes must be selected to convert crude oil into products according to demand. A refinery must also be flexible and be able to change operations as needed, especially if heavier oils are the primary feedstocks. This is accomplished through two basic process concepts: carbon rejection, eg, coking processes, and hydrogen addition, eg, hydroprocesses. However, certain downstream processes, such as catalytic reforming, applied to the product streams do not fit into either of these categories (see CATALYSIS).

In general, when the product is a fraction from crude oil that includes a large number of individual hydrocarbons, the fraction is classified as a refined product. Examples of refined products are gasoline, diesel fuel, heating oils, lubricants, waxes, asphalt, and coke. In contrast, when the product is limited to, perhaps, one or two specific hydrocarbons of high purity, the fraction is classified as a petrochemical product. Examples of petrochemicals are ethylene (qv), propylene (qv), benzene (qv), toluene, and xylene (see BTX PROCESSING).

The application designed for a product requires detailed specifications for various properties; such specifications are set by organizations varying from country to country. For example, in the United States, the American Society for Testing and Materials (ASTM) and the American Petroleum Institute (API) are recognized for establishing specifications on both products and methods for testing. In the United Kingdom, it is the Institute of Petroleum (IP); in Germany, it is Deutsche Institut für Normung (DIN); and in Japan, it is the Ministry of International Trade and Industry (MITI).

Distillation. This is the point at which refining begins and was the first method by which petroleum was refined. Originally, distillation (qv) involved a batch operation in which the still was a cast-iron vessel mounted on brickwork over a fire and the volatile materials were passed through a pipe or gooseneck which led from the top of the still to a condenser. The latter was a coil of pipe, or a "worm" (hence the expression worm end products), immersed in a tank of running water.

Atmospheric Distillation. The petroleum distillation unit in the 1990s brings about a fairly efficient degree of fractionation (separation). The feed to a distillation tower is heated by flow through pipes arranged within a large furnace. The heating unit is known as an atmospheric pipe still heater or pipe still furnace, and the heating unit and fractional distillation tower make up the essential parts of a distillation unit or pipe still.

The pipe still furnace heats the feed to a predetermined temperature, usually a temperature at which a predetermined portion of the feed changes into vapor. The vapor is held under pressure in the pipe in the furnace until it discharges as a foaming stream into the fractional distillation tower. Here the nonvolatile or liquid portion of the feed descends to the bottom of the tower to be pumped away as a bottom nonvolatile product, while the vapors pass up the tower to be fractionated into gas oil, kerosene, and naphtha.

Pipe still furnaces vary greatly and, in contrast to the early units where capacity was usually 31.8–79.5 m³/d (200–500 bbl/d), can now accommodate 3975 m³ (25,000 bbl) or more of crude oil per day. The walls and ceiling are insulated with firebrick and the interior of the furnace is partially divided into two sections: a smaller convection section where the oil first enters the furnace and a larger section fitted with heaters where the oil reaches its highest temperature.

The primary fractions from a distillation unit are equilibrium mixtures and contain some proportion of the lighter constituents characteristic of a lower boiling fraction. The primary fractions are stripped of these constituents (stabilized) before storage or further processing.

Vacuum Distillation. Vacuum distillation evolved as the need arose to separate the less volatile products, such as lubricating oils, from petroleum without subjecting these higher boiling materials to cracking conditions. The boiling point of the heaviest cut obtainable at atmospheric pressure (101.3 kPa = 760 mm Hg) is limited by the temperature (ca 350°C) at which the residue starts to decompose or crack. It is at this point that distillation in a vacuum pipe still is initiated.

Operating conditions for vacuum distillation are usually in the range of 7–13 kPa (50–100 mm Hg). In order to minimize large fluctuations in pressure in the vacuum tower, the units must have a larger diameter than the atmospheric units, ie, greater than 14 meters in diameter. By this means, a heavy gas oil can be obtained as an overhead product at temperatures of about 150°C and lubricating oil cuts can be obtained at 250–350°C. In some designs, the partial pressure of the hydrocarbons is reduced still further by the injection of steam which is added to the column for stripping the nonvolatile constituents in the base of the column.

Azeotropic and Extractive Distillations. Effective as they are for producing various liquid fractions, distillation units generally do not produce specific fractions. In order to accommodate the demand for such products, refineries have incorporated azeotropic distillation and extractive distillation in their operations (see DISTILLATION, AZEOTROPIC AND EXTRACTIVE).

The principle of azeotropic distillation depends on the ability of a chemically dissimilar compound to cause one or both components of a mixture to boil at a temperature other than the one expected. Thus, the addition of a nonindigenous component forms an azeotropic mixture with one of the components of the mixture, thereby lowering the boiling point and facilitating separation by distillation. The separation of components of similar volatility may become economical if an entrainer can be found that effectively changes the relative volatility. It is also desirable that the entrainer be reasonably cheap, stable, nontoxic, and readily recoverable from the components. In practice, it is probably the ready recoverability that limits the application of extractive and azeotropic distillation.

The majority of successful processes are those in which the entrainer and one of the components separate into two liquid phases on cooling if direct recovery by distillation is not feasible. A further restriction in the selection of an azeotropic entrainer is that the boiling point of the entrainer is 10–40°C below that of the components.

Thermal Cracking. When kerosene (lamp oil) was the principal desired product, gasoline was the portion of crude petroleum too volatile to be included in kerosene. The refiners of the 1890s and early 1900s had no use for gasoline and often discarded an accumulation of it. However, as the demand for gasoline and aviation fuel increased with the onset of World War I and the increased production of automobiles during the 1920s, more fuels had to be produced to meet the demand.

The problem of how to produce more of the lower boiling fractions from crude oil was solved in 1913 when cracking units were incorporated into refinery operations and fractions heavier than gasoline were converted into gasoline by thermal decomposition. The use of residua as feedstocks for thermal processes has become economically advantageous because, on the one hand, the end result is the production of lower boiling salable materials, and, on the other, the asphaltic materials in the residua are regarded as the unwanted coke-forming constituents.

Visbreaking. Viscosity breaking (reduction) is a mild cracking operation used to reduce the viscosity of residual fuel oils and residua (8). The process, evolved from the older and now obsolete thermal cracking processes, is classed as mild because the thermal reactions are not allowed to proceed to completion.

Residua are sometimes blended with lighter heating oils to produce fuel oils of acceptable viscosity. By reducing the viscosity of the nonvolatile fraction, visbreaking reduces the amount of the more valuable light heating oil that is required for blending to meet the fuel oil specifications. The process is also used to reduce the pour point of a waxy residue. Visbreaking conditions range from 455–510°C and 345–2070 kPa (50–300 psi) at the heating coil outlet. Liquid-phase cracking takes place under these low severity conditions. In addition to the primary product, fuel oil, material in the gas oil and gasoline boiling range is produced. Gas oil can be used as additional feed for catalytic cracking units or as heating oil.

In the process (Fig. 2), a crude oil residuum is passed through a furnace where it is heated to ca 480°C under an outlet pressure of about 690 kPa (100 psi). The heating coils in the furnace are arranged to provide a soaking section of low heat density, where the charge remains until the visbreaking reactions are completed; subsequently, the cracked products are passed into a flash-distillation chamber. The overhead material from this chamber is then fractionated to produce a low quality gasoline as an overhead product and a light gas oil as bottom. The liquid products from the flash chamber are cooled with a gas oil flux and then sent to a vacuum fractionator. This yields a heavy gas oil distillate and a residual tar of reduced viscosity. A quench oil may also be used to terminate the reactions (2,9).

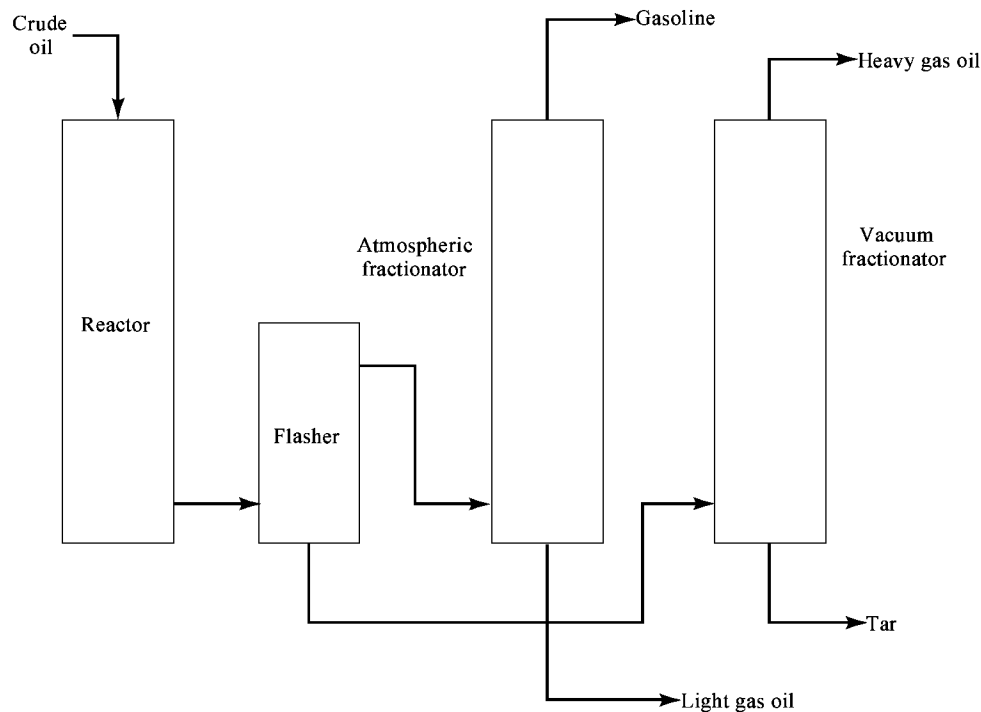


Fig. 2. The visbreaking process.

The main limitation to thermal conversion is that the products can be unstable. Thermal cracking at low pressure gives olefins, particularly in the naphtha fraction; such olefins yield an unstable product that tends to form gum as well as heavier products that form sediments (5).

Coking Processes. Coking is a generic term for a series of thermal processes used for the conversion of nonvolatile heavy feedstocks into lighter, distillable products (10). The feedstock is typically a residuum and the products are gas, naphtha, fuel oil, gas oil, and coke. Gas oil can be the primary product of a coking operation and serves primarily as a feedstock for catalytic cracking units. The coke obtained is usually used as fuel, but specialty uses, such as electrode manufacture and the production of chemicals and metallurgical coke, are also possible, thus increasing the value of the coke.

Delayed coking (Fig. 3) is a semicontinuous process in which the heated charge is transferred to large soaking, or coking, drums, which provide the residence time needed for the cracking reactions to proceed to completion (11,12). The feed to these units is normally a vacuum residuum, although residua from other thermal processes are also used. The feedstock is introduced into the product fractionator. The fractionator bottoms, including a recycle stream of heavy product, are heated in a furnace whose outlet temperature varies from 480–515°C. The heated feedstock then enters one of a pair of coking drums where the cracking reactions continue.

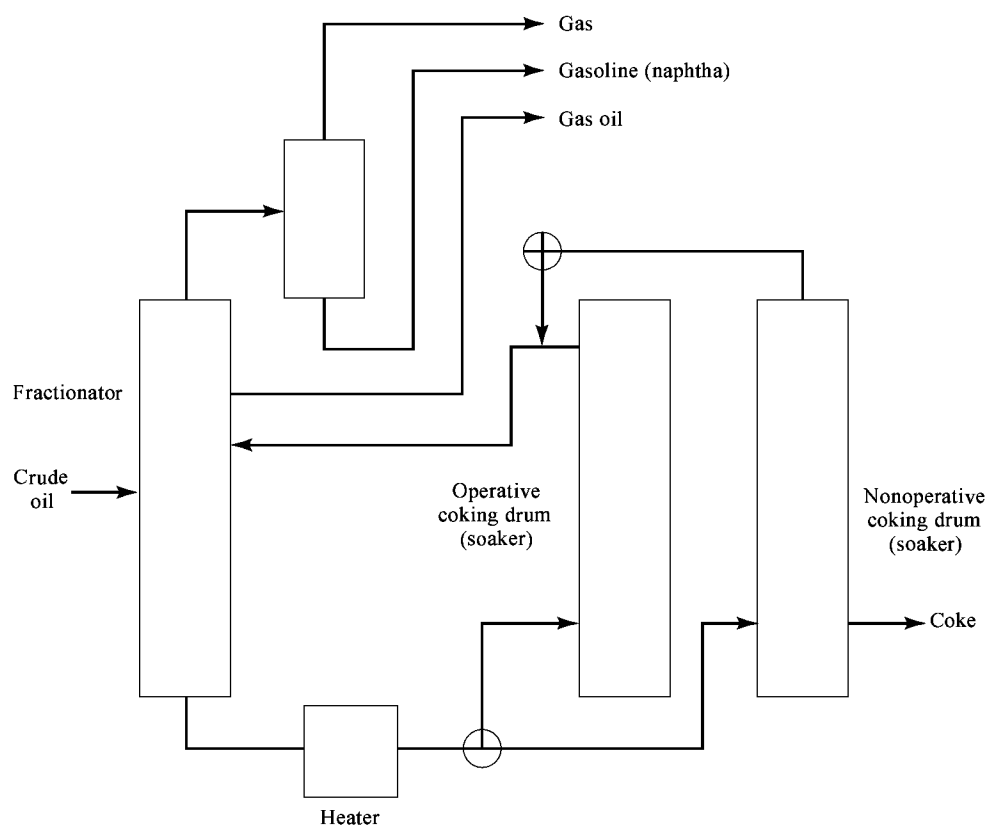


Fig. 3. The delayed coking process.

The cracked products leave as overhead materials, and coke deposits form on the inner surface of the drum. To provide continuous operation, two drums are used; while one drum is on-stream, the one off-stream is being cleaned, steamed, water-cooled, and decoked in the same time interval. The temperature in the coke drum is in the range of 415–450°C with pressures in the range of 103–621 kPa (15–90 psi). Overhead products go to the fractionator, where naphtha and heating oil fractions are recovered. The nonvolatile material is combined with preheated fresh feed and returned to the furnace. The coke drum is usually on stream for about 24 hours before becoming filled with porous coke, after which the coke is removed hydraulically.

Fluid coking (Fig. 4) is a continuous process that uses the fluidized solids technique to convert atmospheric and vacuum residua to more valuable products (12,13). The residuum is converted to coke and overhead products by being sprayed into a fluidized bed of hot, fine coke particles, which permits the coking reactions to be conducted at higher temperatures and shorter contact times than they can be in delayed coking. Moreover, these conditions result in decreased yields of coke; greater quantities of more valuable liquid product are recovered in the fluid coking process.

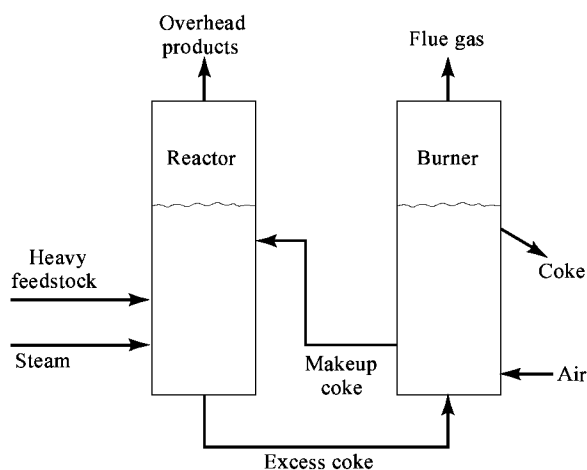


Fig. 4. Fluid coking.

Fluid coking uses two vessels: a reactor and a burner; coke particles are circulated between the two to transfer heat generated by burning a portion of the coke to the reactor. The reactor holds a bed of fluidized coke particles, and steam is introduced at the bottom of the reactor to fluidize the bed.

Flexicoking (Fig. 5), also a continuous process, uses the same configuration as the fluid coker but has a gasification section in which excess coke can be gasified to produce refinery fuel gas. Flexicoking is a process by which excess coke-make is reduced in view of the gradual incursion of the heavier feedstocks into refinery operations. Such feedstocks are notorious for producing high yields of coke (>15% by weight) in thermal and catalytic operations.

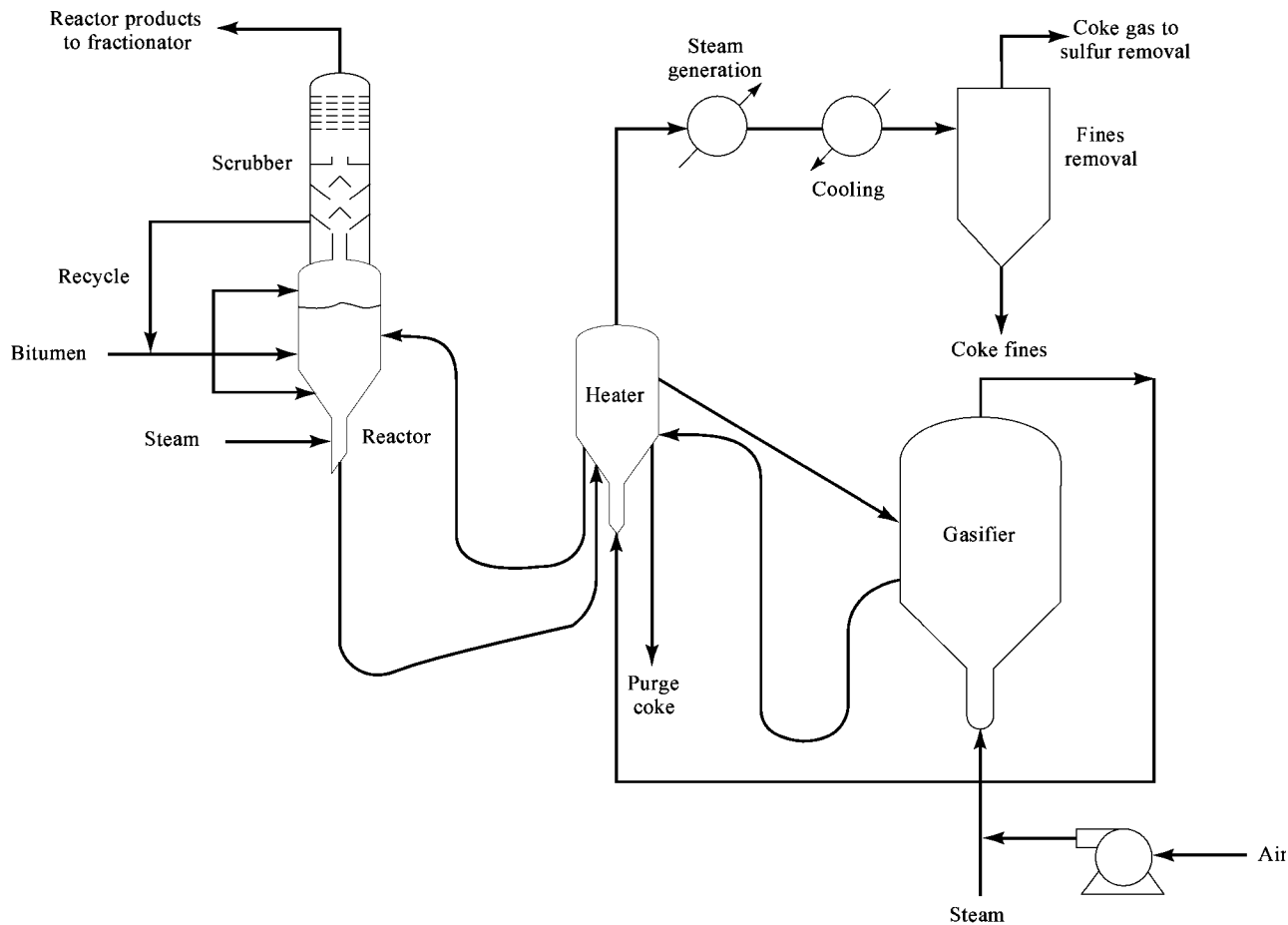


Fig. 5. The flexicoking process.

Catalytic Cracking. Catalytic cracking (Fig. 6), which has progressively supplanted thermal cracking, is the thermal decomposition of petroleum constituents in the presence of a catalyst (14,15). The acid catalysts first used in catalytic cracking were designated low alumina catalysts; amorphous solids composed of approximately 87% silica, SiO_2 , and 13% alumina, Al_2O_3 . Later, high alumina catalysts containing 25% alumina and 75% silica were used. However, this type of catalyst has largely been replaced by catalysts containing crystalline aluminosilicates (zeolites) or molecular sieves (qv) (16–18).

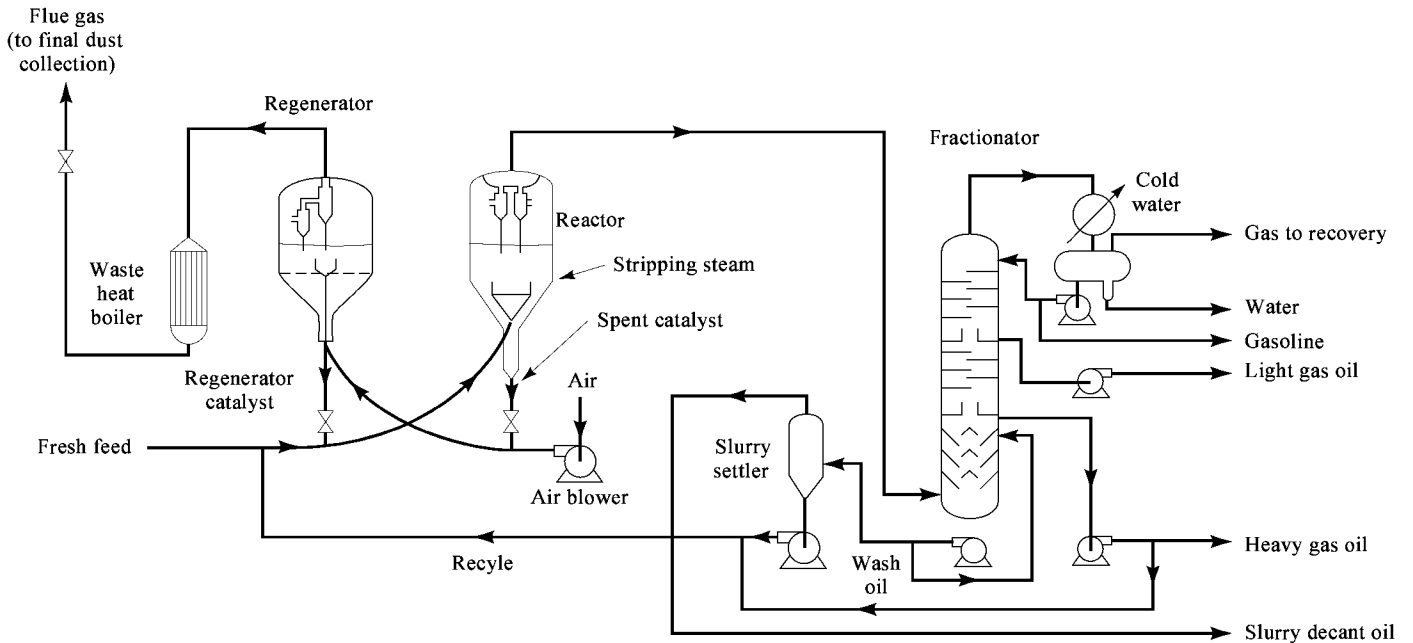


Fig. 6. Fluid-bed catalytic cracking with product separation.

The catalyst is employed in bead, pellet, or microspherical form and can be used as a fixed bed, moving bed, or fluid bed. The fixed-bed process was the first process used commercially and employs a static bed of catalyst in several reactors, which allows a continuous flow of feedstock to be maintained. The cycle of operations consists of (1) the flow of feedstock through the catalyst bed; (2) the discontinuance of feedstock flow and removal of coke from the catalyst by burning; and (3) the insertion of the reactor back on-stream. The moving-bed process uses a reaction vessel, in which cracking takes place, and a kiln, in which the spent catalyst is regenerated and catalyst movement between the vessels is provided by various means.

Table 2. General Process Characteristics for Hydroprocessing Various Feedstocks

Process	Naphtha	Atmospheric gas oil	Vacuum gas oil	Residuum
			oil	

hydrocracking	+	+	+	+
aromatics removal		+	+	
sulfur removal	+		+	
nitrogen removal	+		+	+
metals removal			+	+
coke mitigation				+
<i>n</i> -paraffins removal		+		
olefins removal	+			

Hydroprocessing. In this group of refining processes, which includes hydrotreating and hydrocracking, the feedstock is heated with hydrogen at high temperature and under pressure. The outcome is the conversion of a variety of feedstocks to a range of products (Table 2) (2,6,9,19). The purpose of hydroprocessing is (1) to improve existing petroleum products or develop new products or uses; (2) to convert inferior or low grade materials into valuable products; and (3) to transform near-solid residua to liquid fuels. Products are as follows: from naphtha, reformed feedstock and liquefied petroleum gas (LPG); from atmospheric gas oil, diesel and jet fuel, petrochemical feedstock, and naphtha; from vacuum gas oil, catalytic cracker feedstock, kerosene, diesel and jet fuel, naphtha, LPG, and lubricating oil; and from residuum, catalytic cracker and coker feedstock, diesel fuel, and others.

Hydroprocesses for the conversion of petroleum and petroleum products can be classified as destructive or nondestructive. The former (hydrogenolysis and hydrocracking) is characterized by the rupture of carbon–carbon bonds and is accompanied by hydrogen saturation of the fragments to produce lower boiling products. Such treatment requires rather high temperatures and high hydrogen pressures, the latter to minimize coke formation.

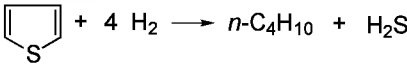
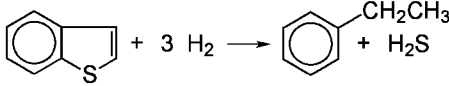
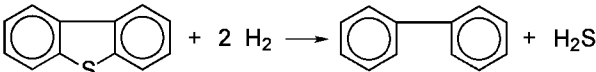
Hydrogenolysis is analogous to hydrolysis and ammonolysis, which involve the cleavage of a bond induced by the action of water and ammonia, respectively. Chemical bonds that are broken by hydrogenolysis reactions include carbon–carbon, carbon–oxygen, carbon–sulfur, and carbon–nitrogen bonds. An example of hydrogenolysis is the hydrodealkylation of toluene to form benzene and methane:



On the other hand, nondestructive, or simple hydrogenation is generally used for the purpose of improving product quality without appreciable alteration of the boiling range. Examples are the removal of various sulfur compounds (Table 3) which would otherwise have an adverse effect on product quality. Treatment under such mild conditions is often referred to as hydrotreating or hydrofining and is essentially a means of eliminating, in addition to sulfur, nitrogen and oxygen as ammonia and water, respectively.

Hydrocracking. Hydrocracking is a catalytic process (>350°C) in which hydrogenation accompanies cracking (20). Relatively high pressures, 6,900–21,000 kPa (1000–3000 psi), are employed and the overall result is the conversion of the feedstock to lower boiling products. Another attractive feature of hydrocracking is the low yield of gaseous components, such as methane, ethane, and propane, which are less desirable than the gasoline components. Essentially all the initial reactions of catalytic cracking occur, but some of the secondary reactions are inhibited or stopped by the presence of hydrogen.

Table 3. Sulfur Removal from Organic Compounds by Hydrotreating

Name	Typical reaction
thiols (mercaptans)	$\text{R-SH} + \text{H}_2 \rightarrow \text{RH} + \text{H}_2\text{S}$
disulfides	$\text{R-S-S-R}' + 3\text{H}_2 \rightarrow \text{RH} + \text{R}'\text{H} + 2\text{H}_2\text{S}$
sulfides	$\text{R-S-R}' + 2\text{H}_2 \rightarrow \text{RH} + \text{R}'\text{H} + \text{H}_2\text{S}$
thiophenes	
benzothiophenes	
dibenzothiophenes	

In the first, pretreating stage of a hydrocracking process, the main purpose is to convert organic nitrogen compounds and organic sulfur in the feedstock to hydrocarbons and to ammonia and hydrogen sulfide by hydrogenation and mild hydrocracking. The purpose is to reduce the organic nitrogen and sulfur compounds to low levels (<50 ppm). Typical conditions are 340–390°C, 10,350–17,250 kPa (1500–2500 psi), and a catalyst contact time of 0.5–1.5 h; up to 1.5 wt % hydrogen is absorbed, partly by conversion of the nitrogen compounds, but chiefly by aromatic compounds that are hydrogenated.

This first stage is usually carried out with a bifunctional catalyst containing hydrogenation promoters, eg, nickel and tungsten or molybdenum sulfides, on an acidic support. The metal sulfides hydrogenate aromatics and nitrogen compounds, and retard deposition of carbonaceous deposits; the acidic support accelerates nitrogen removal as ammonia by breaking carbon–nitrogen bonds. The catalyst is generally used as 0.32 × 0.32 cm or 0.16 × 0.32 cm pellets, as well as spheres or other shapes.

Most of the hydrocracking is accomplished in the second stage. Hydrogen sulfide, ammonia, and low boiling products are usually removed from the first-stage product; the remaining oil, which is low in nitrogen–sulfur compounds, is passed over the second-stage catalyst. Some catalyst systems do not require the removal of adsorbed compounds between stages. In the second stage, typical conditions are 300–370°C, 7,000–17,250 kPa (1000–2500 psi) hydrogen pressure, and 0.5–1.5-h contact time; 1–1.5 wt % hydrogen may be absorbed. Conversion to gasoline or jet fuel is seldom complete in one contact with the catalyst, so the lighter oils are removed by distillation of the products, and the heavier, high boiling product is combined with fresh feed and recycled over the catalyst until it is completely converted.

The catalyst for the second stage is also a bifunctional catalyst containing hydrogenating and acidic components. Metals such as nickel, molybdenum, tungsten, or palladium are used in various combinations and dispersed on solid acidic supports such as synthetic amorphous or crystalline silica–alumina, eg, zeolites. These supports contain strongly acidic sites and sometimes are enhanced by the incorporation of a small amount of fluorine.

A long period of operation (eg, >3 yr) between catalyst regeneration is desirable; this is achieved by keeping a low nitrogen content in the feed and avoiding high temperatures and high end-point feedstock, which leads to excess cracking and consequent deposition of coke on the catalyst. Feedstock conversion is the key insofar as the conversion dictates the temperature employed. When activity of the catalyst has decreased, it can often be restored by controlled burning of the coke (see CATALYSTS, REGENERATION).

Hydrotreating. This catalytic process converts sulfur- and or nitrogen-containing hydrocarbons into low sulfur low nitrogen liquids, hydrogen sulfide, and ammonia (21). A wide variety of metals are active hydrogenation catalysts; those of most interest are nickel, palladium, platinum, cobalt, and iron. Special preparations of the first three are active at room temperature and atmospheric pressure. The metallic catalysts are easily poisoned by sulfur- or arsenic-containing compounds, and even by other metals. To avoid such poisoning, less effective but more resistant metal oxides or sulfides are frequently employed, generally those of tungsten, cobalt, or molybdenum. Alternatively, catalyst poisoning can be minimized by mild hydrogenation to remove nitrogen, oxygen, and sulfur from feedstocks in the presence of more resistant catalysts, such as cobalt–molybdenum–alumina, Co–Mo–Al₂O₃.

The process temperature affects the rate and the extent of hydrogenation as it does any chemical reaction. Practically every hydrogenation reaction can be reversed by increasing temperature. If a second functional group is present, high temperatures often lead to the loss of selectivity and, therefore, loss of desired product yield. As a practical measure, hydrogenation is carried out at as low a temperature as possible which is still compatible with a satisfactory reaction rate.

Hydrotreating is carried out by charging the feed to the reactor together with hydrogen at 300–345°C; the hydrogen pressures are about 3450–6900 kPa (500–1000 psi). The reaction generally takes place in the vapor phase but, depending on the application, can also be a mixed-phase reaction.

After passing through the reactor, the treated oil is cooled and separated from the excess hydrogen recycled through the reactor. The treated oil is pumped to a stripper tower where hydrogen sulfide, formed by the hydrogenation reaction, is removed by steam or by hydrocarbon vapor via reboiling, and the finished product leaves the bottom of the stripper tower. The catalyst can be regenerated *in situ* and ultimately be replaced after several regenerations.

Reforming. When the demand for higher octane gasolines increased during the early 1930s, attention was directed to ways and means of improving the octane number of fractions within the boiling range of gasoline. Straight-run (distilled) gasolines frequently had low octane numbers, and any process that could improve the octane numbers would aid in meeting the demand for higher octane number gasoline. Such a process, called thermal reforming, was developed and used widely, but to a much lesser extent than thermal cracking. Thermal reforming was a natural development from older thermal cracking processes; cracking converts heavier oils into gasoline whereas reforming converts (reforms) gasolines into higher octane gasolines. The equipment for thermal reforming is essentially the same as for thermal cracking, but higher temperatures are used in the former.

Thermal reforming, less effective and less economical than catalytic processes, has been largely supplanted. Like thermal reforming, catalytic reforming converts low octane gasolines into high octane gasolines, ie, reformate. Whereas thermal reforming produces reformate having research octane numbers in the 65–80 range, depending on the yield, catalytic reforming produces reformate having octane numbers on the order of 90–105. Catalytic reforming is conducted in the presence of hydrogen over hydrogenation–dehydrogenation catalysts, eg, in the platforming process (22). Catalytic reformer feeds are saturated, ie, not olefinic, materials. Catalytic cracker naphtha and hydrocracker naphtha that contains substantial quantities of naphthenes are also suitable reformer feedstocks.

Dehydrogenation is a main chemical reaction in catalytic reforming, and hydrogen gas is consequently produced in large quantities. Hydrogen is recycled through the reactors where the reforming takes place to provide the atmosphere necessary for the chemical reactions, and also prevents carbon from being deposited on the catalyst, thus extending its operating life. Because an excess of hydrogen above whatever is consumed in the process is produced, catalytic reforming processes are unique in that they are the only petroleum refinery processes to produce hydrogen as a by-product.

Catalytic reforming usually is carried out by feeding a naphtha (after pretreating with hydrogen if necessary to remove nitrogen and sulfur compounds) and hydrogen mixture to a furnace where the mixture is heated to the desired temperature (450–520°C) and then passed through fixed-bed catalytic reactors at hydrogen pressures of 350–2700 kPa (50–400 psi). Normally, several reactors are used in series and heaters are located between adjoining reactors in order to compensate for the endothermic reactions taking place.

The composition of a reforming catalyst is dictated by the composition of the feedstock and the desired reformate. The catalysts used are principally platinum or platinum–rhenium on an alumina base. The purpose of platinum on the catalyst is to promote dehydrogenation and hydrogenation reactions. Nonplatinum catalysts are used in regenerative processes for feedstocks containing sulfur, although pretreatment (hydrodesulfurization) may permit platinum catalysts to be employed.

Isomerization. Isomerization is used with the objective of providing additional feedstock for alkylation units (isobutane) or high octane fractions for gasoline blending (pentane and hexane) (23). The latter application is useful in the production of reformulated gasoline by increasing the octane number while converting or removing benzene (24,25). Isobutane is also used for the synthesis of methyl *t*-butyl ether (MTBE), an additive that maintains the octane ratings of gasoline in the absence of added lead.

Initially, aluminum chloride was the catalyst used to isomerize butane, pentane, and hexane. Since then, supported metal catalysts have been developed for use in high temperature processes that operate at 370–480°C and 2070–5170 kPa (300–750 psi), whereas aluminum chloride and hydrogen chloride are universally used for the low temperature processes.

Nonregenerable aluminum chloride catalyst is employed with various carriers in a fixed-bed or liquid contactor. Platinum or other metal catalyst processes that utilize fixed-bed operation can be either regenerable or nonregenerable. The reaction conditions vary widely, between 40–480°C and 1035–6900 kPa (150–1000 psi), depending on the particular process and feedstock.

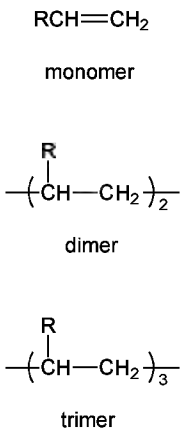
Alkylation. The combination of olefins with paraffins to form higher isoparaffins is termed alkylation (qv). Alkylate is a desirable blendstock because it has a relatively high octane number and serves to dilute the total aromatics content. Reduction of the olefins in gasoline blendstocks by alkylation also reduces tail pipe emissions. In refinery practice, butylenes are routinely alkylated by reaction with isobutane to produce isobutane–octane (26). In some plants, propylene and or pentylenes (amylene) are also alkylated (27).

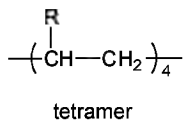
Alkylate is composed of a mixture of isoparaffins whose octane numbers vary with the olefins from which they were made. Butylenes produce the highest octane numbers, propylene the lowest, and amylene (pentylenes) the intermediate values. All alkylates, however, have high (>87) octane numbers that make them particularly valuable.

Propylene, butylenes, or amylene are combined with isobutane in the presence of an acid catalyst, eg, sulfuric acid or hydrofluoric acid, at low temperatures (1–40°C) and pressures, 102–1035 kPa (1–10 atm). Sulfuric acid or hydrogen fluoride are the catalysts used commercially in refineries. The acid is pumped through the reactor and forms an emulsion with reactants, and the emulsion is maintained at 50° acid. The rate of deactivation varies with the feed and isobutane charge rate. Butene feeds cause less acid consumption than the propylene feeds.

Polymerization. In the petroleum industry, polymerization is the process by which olefin gases are converted to higher molecular weight liquid products which may be suitable for gasoline (polymer gasoline) or other liquid fuels.

The feedstock, usually consisting of propylene and butylenes (various isomers of C₄H₈) from cracking processes, may even consist of selective olefins for dimer, trimer, or tetramer production:





The molecular size of the product is limited insofar as the reaction is terminated at the dimer or trimer stage. Thus the process is more properly termed oligomerization. The four- to twelve-carbon compounds required as the constituents of liquid fuels are the prime products.

Thermal polymerization is not as effective as catalytic polymerization but has the advantage that it can be used to polymerize saturated materials that cannot be induced to react by catalysts. The process consists of the vapor-phase cracking of, for example, propane and butane, followed by prolonged periods at high temperature (510–595°C) for the reactions to proceed to near completion. Olefins can also be conveniently polymerized by means of an acid catalyst. Thus, the treated olefin-rich feed stream is contacted with a catalyst, such as sulfuric acid, copper pyrophosphate, or phosphoric acid, at 150–220°C and 1035–8275 kPa (150–1200 psi), depending on feedstock and product requirement.

Phosphates are the principal catalysts used in polymerization units; the commercially used catalysts are liquid phosphoric acid, phosphoric acid on kieselguhr, copper pyrophosphate pellets, and phosphoric acid film on quartz. The last is the least active and has the disadvantage that carbonaceous deposits must occasionally be burned off the support. Compared to other processes, the one using liquid phosphoric acid catalyst is far more responsive to attempts to raise production by increasing temperature.

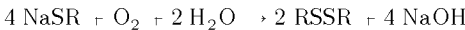
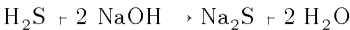
Treating

Since the original crude oils contain some sulfur compounds, the resulting products and gasolines also contain sulfur compounds, including hydrogen sulfide, mercaptans, sulfides, disulfides, and thiophenes. The processes used to sweeten, ie, desulfurize, the products depend on the type and amount of the sulfur compounds present and the specifications of the finished gasoline or other stocks.

Hydrotreating is the most widely practiced treating process for all types of petroleum products. However, there are other treating processes suitable for the removal of mercaptans and hydrogen sulfide; such processes are necessary and are performed as part of the product improvement and finishing procedures. For example, mercaptan, RSH, removal is achieved by using regenerative solution processes, in which the treatment solutions are regenerated rather than discarded. Mercaptan conversion is essentially a process of oxidation to disulfides, RSSR, by lead sulfide treatment, copper chloride–oxygen treatment, sodium hypochlorite treatment, or oxygen treatment with a chelated cobalt catalyst in either a caustic solution or fixed bed.

Hydrogen sulfide, H₂S, is removed by a variety of processes, of which one is a regenerative solution process using aqueous solutions of sodium hydroxide, NaOH, calcium hydroxide, Ca(OH)₂, sodium phosphate, Na₃PO₄, and sodium carbonate, Na₂CO₃.

Alkali Treatment. Caustic washing is the treatment of materials, usually products from petroleum refining, with solutions of caustic soda. The process consists of mixing a water solution of lye (sodium hydroxide or caustic soda) with a petroleum fraction. The treatment is carried out as soon as possible after the petroleum fraction is distilled, since contact with air forms free sulfur, which is corrosive and difficult to remove. The lye reacts either with any hydrogen sulfide present to form sodium sulfide, which is soluble in water, or with mercaptans, followed by oxidation, to form the less noxious disulfides.



Nonregenerative caustic treatment is generally economically applied when the contaminating materials are low in concentration and waste disposal is not a problem. However, the use of nonregenerative systems is on the decline because of the frequently occurring waste disposal problems that arise from environmental considerations and because of the availability of numerous other processes that can effect more complete removal of contaminating materials.

Steam-regenerative caustic treatment is directed toward the removal of mercaptans from such products as gasoline and low boiling solvents (naphtha). The caustic is regenerated by steam blowing in a stripping tower. The nature and concentration of the mercaptans to be removed dictate the quantity and temperature of the process. However, the caustic gradually deteriorates because of the accumulation of material that cannot be removed by stripping; the caustic quality must be maintained by either continuous or intermittent discard and or replacement of a minimum amount of the operating solution.

Acid Treatment. The treatment of petroleum products with acids has been in use for a considerable time in the petroleum industry. Various acids such as hydrofluoric acid, hydrochloric acid, nitric acid, and phosphoric acid have been used in addition to the most commonly used sulfuric acid, but in most instances there is little advantage in using any acid other than sulfuric.

Sulfuric acid also has been employed for refining kerosene distillates and lubricating oil stocks. Although a greater part of the acid-treating processes has been superseded by other processes, acid treating has continued to some extent; it is used for desulfurizing high boiling fractions of cracked gasoline distillates, for refining paraffinic kerosene, for manufacturing low cost lubricating oils, and for making specialty products such as insecticides, pharmaceutical oils, and insulating oils.

Clay Treatment. The original method of clay treating was to percolate a petroleum fraction through a tower containing coarse clay pellets. As the clay adsorbed impurities from the petroleum fraction, the clay became less effective. The activity of the clay was periodically restored by removing it from the tower and burning the adsorbed material under carefully controlled conditions so as not to sinter the clay. The percolation method of clay treating was widely used for lubricating oils, but has been largely replaced by clay contacting.

However, this use of clay treating has been superseded by other processes; in particular, by the use of inhibitors, which, when added in small amounts to gasoline, can prevent gums from forming. Nevertheless, clay treating is still used as a finishing step in the manufacture of lubricating oils and waxes. The clay removes traces of asphaltic materials and other compounds that give oils and waxes unwanted odors and colors.

Solvent Treatment. Solvent processes can be divided into two main categories, solvent extraction and solvent dewaxing. The solvent used in the extraction processes include propane and cresylic acid, 2,2'-dichlorodiethyl ether, phenol (qv), furfural, sulfur dioxide, benzene, and nitrobenzene. In the dewaxing process (28), the principal solvents are benzene, methyl ethyl ketone, methyl isobutyl ketone, propane, petroleum naphtha, ethylene dichloride, methylene chloride, sulfur dioxide, and *N*-methylpyrrolidinone.

The early developments of solvent processing were concerned with the lubricating oil end of the crude. Solvent extraction processes are applied to many useful separations in the purification of gasoline, kerosene, diesel fuel, and other oils. In addition, solvent extraction can replace fractionation in many separation processes in the refinery. For example, propane deasphalting (Fig. 7) has replaced, to some extent, vacuum distillation as a means of removing asphalt from reduced crude oils.

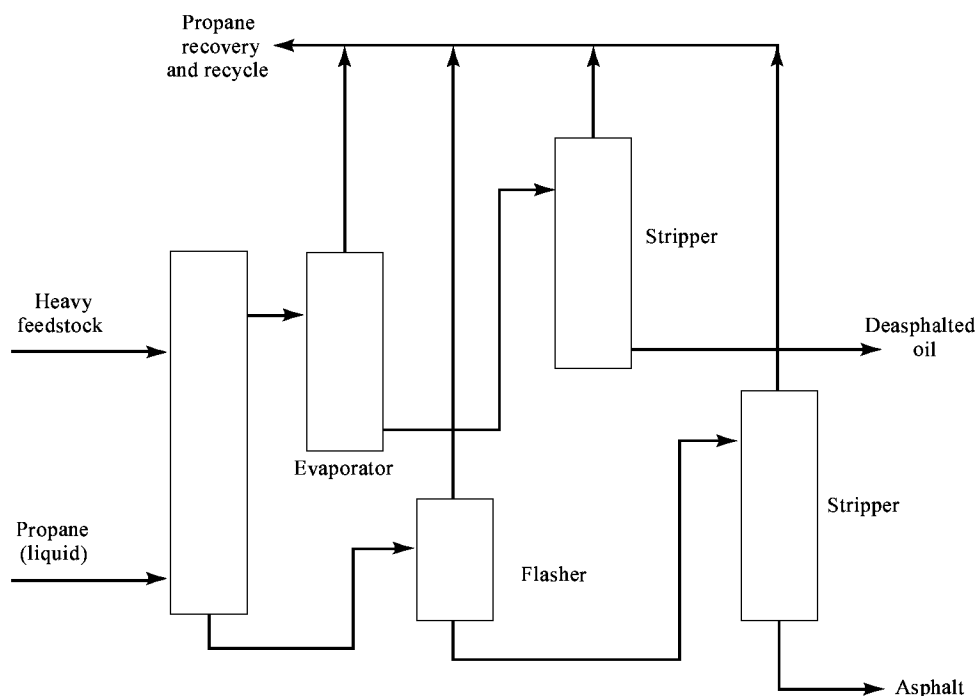


Fig. 7. Propane deasphalting.

Gas Processing

The gas streams produced during petroleum refining usually contain many noxious constituents that have an adverse effect on the use of the gas for other purposes, eg, as a fuel or as a petrochemical feedstock, and some degree of cleaning is required (29).

Gas purification processes fall into three categories: the removal of gaseous impurities, the removal of particulate impurities, and ultrafine cleaning. The extra expense of the last process is only justified by the nature of the subsequent operations or the need to produce a pure gas stream. Because there are many variables in gas treating, several factors must be considered: (1) the types and concentrations of contaminants in the gas; (2) the degree of contaminant removal desired; (3) the selectivity of acid gas removal required; (4) the temperature, pressure, volume, and composition of the gas to be processed; (5) the carbon dioxide-to-hydrogen sulfide ratio in the gas; and (6) the desirability of sulfur recovery on account of process economics or environmental issues.

Process selectivity indicates the preference with which the process removes one acid gas component relative to or in preference to another. For example, some processes remove both hydrogen sulfide and carbon dioxide, whereas other processes are designed to remove hydrogen sulfide only. Thus it is important to consider the process selectivity for hydrogen sulfide removal compared to carbon dioxide removal, ie, the carbon dioxide-to-hydrogen sulfide ratio in the natural gas, in order to ensure minimal concentrations of these components in the product.

One of the principal aspects of refinery gas cleanup is the removal of acid gas constituents, ie, carbon dioxide, CO₂, and hydrogen sulfide, H₂S. Treatment of natural gas to remove the acid gas constituents is most often accomplished by contacting the natural gas with an alkaline solution. The most commonly used treating solutions are aqueous solutions of the ethanolamines or alkali carbonates. There are several hydrogen sulfide removal processes (29), most of which are followed by a Claus plant that produces elemental sulfur from the hydrogen sulfide.

Products

The constant demand for products such as liquid fuels is the main driving force behind the petroleum industry (7,30). In fact, it is the changes in product demand that have been largely responsible for the evolution of the industry.

Liquefied Petroleum Gas (LPG). Certain specific hydrocarbons, such as propane, butane, pentane, and their mixtures, exist in the gaseous state under atmospheric ambient conditions but can be converted to the liquid state under conditions of moderate pressure at ambient temperature. This is termed liquefied petroleum gas (LPG). Liquefied petroleum gas (qv) is a refinery product and the individual constituents, or light ends (Table 4), are produced during a variety of refining operations.

The presence of ethane in liquefied petroleum gas must be avoided because of the inability of this lighter hydrocarbon to liquefy under pressure at ambient temperature and its tendency to register abnormally high pressures in the LPG containers. On the other hand, the presence of pentane in liquefied petroleum gas must also be avoided because this particular hydrocarbon, a liquid at ambient temperatures and pressures, can separate in the liquid state in the gas lines.

Liquefied petroleum gases precipitate asphaltic and resinous materials from crude residues while the lubricating oil constituents remain in solution. Although all liquefied gases possess this property to some extent, propane and butane are used to deasphalt residual lubricating oils because of their relative low cost and their ease of separation from lubricating oils.

Table 4. Constituents of Light Ends

Hydrocarbon	Carbon atoms	Molecular weight	Boiling point, °C	Uses
methane	1	16	−182	fuel gas
ethane	2	30	−89	fuel gas
ethylene	2	28	104	fuel gas, petrochemicals
propane	3	44	42	fuel gas, LPG
propylene	3	42	48	fuel gas, petrochemicals, polymer gasoline
isobutane	4	58	12	alkylate, motor gasoline
<i>n</i> -butane	4	58	1	motor gasoline
isobutylene	4	56	7	synthetic rubber and chemicals, polymer gasoline, alkylate, motor gasoline
butylene-1 ^a	4	56	6	synthetic rubber and chemicals,
butylene-2 ^a	4	56	1	alkylate, polymer gasoline, motor gasoline

isopentane	5	72	28	motor and aviation gasolines
<i>n</i> -pentane	5	72	36	motor and aviation gasolines
pentylenes	5	70	30	motor gasolines
isohexane	6	86	61	motor and aviation gasolines
<i>n</i> -hexane	6	86	69	motor and aviation gasolines

^a Numbers refer to the positions of the double bond; for example, butylene-1 (or butene-1 or but-1-ene) is CH₃CH₂CH=CH₂ and butylene-2 (or butene-2 or but-2-ene) is CH₃CH=CHCH₃.

Gasoline. The naphtha fraction from crude oil distillation is ultimately used to make gasoline. The two streams are isolated early in the refining scheme so that each can be refined separately for optimum blending in order to achieve the required specifications (see GASOLINE AND OTHER MOTOR FUELS).

Gasoline is a complex mixture of hydrocarbons that boils below 200°C. The hydrocarbon constituents in this boiling range are those that have four to twelve carbon atoms in their molecular structure. Gasolines can vary widely in composition, even those having the same octane number can be quite different (30). Because of the differences in composition of the various gasolines, gasoline blending is necessary. The physical process of blending the components is simple, but determination of how much of each component to include in a blend is much more difficult. The operation is carried out by simultaneously pumping all the components of a gasoline blend into a pipeline that leads to the gasoline storage, and the pumps must be set to deliver automatically the proper proportion of each component. Sophisticated instrumentation is employed to achieve the desired blends.

Aviation gasolines, usually used in light aircraft and older civil aircraft, have narrower boiling ranges (38–170°C) than automobile gasolines (0–200°C). In winter, the extreme temperatures in the northern climates allow a certain amount of butane to be dissolved in the gasoline to facilitate vaporization. The narrower boiling range ensures better distribution of the vaporized fuel through the more complicated induction systems of aircraft engines. Aircraft operates at altitudes where the prevailing pressure is less than the pressure at the surface of the earth, eg, at 5334 m, 52 kPa (7.5 psi) compared to 101.3 kPa (14.7 psi). As a result, the vapor pressure of aviation gasolines, which is a function of the fuel's boiling range, must be limited to reduce boiling in the tanks, fuel lines, and carburetors (see AVIATION AND OTHER GAS TURBINE FUELS).

Another condition to keep gasoline engines running smoothly is to allow the fuel–air mixture to start burning at a precise time in the combustion cycle. An electrical spark starts the ignition. The remainder of the fuel–air mixture should be consumed by a flame front moving out from the initial spark. Under certain conditions, a portion of the fuel–air mixture ignites spontaneously instead of waiting for the flame front from the carefully timed spark. The extra pressure pulses resulting from spontaneous combustion are usually audible above the normal sounds of a running engine and give rise to the phenomenon called knock, of which pinging and rumbling are special attributes. However, knocks are undesirable because they waste some of the available power of an otherwise smooth-running engine.

Octane number is a measure of a fuel's ability to avoid knocking. The octane number of a gasoline is determined in a special single-cylinder engine where various combustion conditions can be controlled. The test engine is adjusted to give trace knock from the fuel to be rated. Various mixtures of isooctane (2,2,4-trimethyl pentane) and normal heptane are then used to find the ratio of the two reference fuels that produce the same intensity of knock as that by the unknown fuel.

By defining isooctane as having an octane number of 100 and *n*-heptane as having an octane number of 0, the volumetric percentage of isooctane in heptane that matches the knock from the unknown fuel can be calculated as the octane number of the fuel. For example, 90 vol % isooctane and 10 vol % normal heptane produce a 90-octane-number reference fuel.

Two kinds of octane number ratings are specified, although other methods are often used for engine and fuel development. Both methods use the same reference fuels and essentially the same test engine, but engine operating conditions are different. In one test, called the research method, the spark advance is fixed, the air inlet temperature is 125°F (~ 52°C), and engine speed is 600 r min. The other, called the motor method, uses variable spark timing, a higher mixture temperature of 300°F (~ 149°C), and a faster engine speed of 900 r min. The more severe conditions of the motor method have a greater influence on commercial blends than they do on the reference fuels. Thus, the motor octane number of a commercial blend tends to be lower than the research octane number. Common practice is to label gasoline with an arithmetic average of both ratings, abbreviated as (R + M)/2, and often referred to as road octane number.

Reformulated gasoline is believed to be the answer to many environmental issues that arise from the use of automobiles and there has been a serious effort to produce reformulated gasoline components (Table 5) from a variety of processes (25). However, reformulation may increase gasoline consumption, when in fact the converse is preferable (24). It has also been claimed that methyl *t*-butyl ether (MTBE), an additive that maintains the octane ratings of gasoline in the absence of added lead, can reduce the emissions of unburned hydrocarbons during gasoline use through more efficient combustion of the hydrocarbons (31). However, the ether, MTBE, is believed to have an adverse effect insofar as it appears that aldehyde emissions are increased (24).

Table 5. Refining Technologies for Producing Reformulated Gasoline Constituents

Technology	Objective
catalytic reformer prefractionation	reduce benzene
reformate fractionation	reduce benzene
isomerization	increase octane
aromatics saturation	reduce total aromatics
catalytic reforming	oxygenates for octane enhancement
MTBE synthesis	provide oxygenates
isobutane dehydrogenation	feedstock for oxygenate synthesis
catalytic cracker naphtha fractionation	increase alkylate; increase oxygenates; reduce olefins and sulfur
feedstock hydrotreating	reduce sulfur

Solvents. Petroleum naphtha is a generic term applied to refined, partly refined, or unrefined petroleum products. Naphthas are prepared by any of several methods, including fractionation of distillates or even crude petroleum, solvent extraction, hydrocracking of distillates, polymerization of unsaturated (olefinic) compounds, and alkylation processes. Naphtha can also be a combination of product streams from more than one of these processes.

The main uses of petroleum naphtha fall into the general areas of solvents (diluents) for paints, etc, dry-cleaning solvents, solvents for cutback asphalt, solvents in rubber industry, and solvents for industrial extraction processes. Turpentine, the older, more conventional solvent for paints, has been almost completely replaced by the cheaper and more abundant petroleum naphtha.

Kerosene. Kerosene, also called kerosine, originated as a straight-run (distilled) petroleum fraction that boiled over the temperature range of 205– 260°C. In the early days of petroleum refining, some crude oils contained kerosene fractions of high quality, but other crude oils, such as those having a high proportion of asphaltic materials, had to be thoroughly refined to remove aromatics and sulfur compounds before a satisfactory kerosene fraction could be obtained.

Kerosene is believed to be composed chiefly of hydrocarbons containing twelve to fifteen carbon atoms per molecule. Low proportions of aromatic and unsaturated hydrocarbons are desirable to maintain the lowest possible level of smoke during burning. Although some aromatics may occur within the boiling range assigned to kerosene, excessive amounts can be removed by extraction.

The significance of the total sulfur content of kerosene varies greatly with the type of oil and the use to which it is put. Sulfur content is of great importance when the kerosene to be burned produces sulfur oxides, which are of environmental concern. The color of kerosene is of little significance but a product darker than usual may have resulted from contamination or aging; in fact, a color darker than specified may be considered by some users as unsatisfactory. Kerosene, because of its use as a burning oil, must be free of aromatic and unsaturated hydrocarbons; the desirable constituents of kerosene are saturated hydrocarbons.

Diesel fuel, jet fuel, kerosene (range oil), no. 1 fuel oil, no. 2 fuel oil, and diesel fuel are all popular distillate products coming from the kerosene fraction of petroleum. One grade of jet fuel uses the heavy naphtha fraction, but the kerosene fraction supplies the more popular and heavier grade of jet fuel, as well as smaller amounts which are sold as burner fuel (range oil) or no. 1 heating oil. Some heating oil (generally no. 2 heating oil) and diesel fuel are similar and can sometimes substitute for each other. The home heating oil is intended to be burned with a furnace for space heating. The diesel fuel is intended for compression-ignition engines.

The cetane number of a diesel fuel is a number that indicates the ability of a diesel engine fuel to ignite quickly, and burn smoothly, after being injected into the cylinder. In high speed diesel engines, a fuel with a long ignition delay tends to produce rough operation. The cetane number should not be confused with the cetene number, an obsolete designation for the starting and running quality of diesel fuel that uses cetene, C₁ H₃₀, as the reference fuel. The cetene number has been replaced by the cetane number, a scale based on the ignition characteristics of two well-defined hydrocarbons, cetane, *n*-hexadecane, and 2,3,4,5,6,7,8-heptamethylnonane.

Cetane has a short delay period during ignition and is assigned a cetane number of 100; heptamethylnonane has a long delay period and has a cetane number of 15. Just as the octane number is meaningful for automobile fuels, the cetane number is a means of determining the ignition quality of diesel fuels and is equivalent to the percentage by volume of cetane, in the blend with heptamethylnonane, that matches the ignition quality of the test fuel. The cetane number of diesel fuel usually falls into the 30–60 range; a high cetane number is an indication of the potential for easy starting and smooth operation of the engine.

Other methods are also available for the estimation of diesel fuel quality. For example, the diesel index (DI) is defined by the relation $DI = (A^{\circ}F \times ^{\circ}API) / 100$, where A °F is the aniline point in degrees Fahrenheit and °API is the American Petroleum Institute gravity. A high aniline point corresponds to a high proportion of paraffins in the diesel fuel; such a fuel has a high diesel index and, therefore, a high cetane number.

Fuel Oil. Fuel oil is classified in several ways, but generally into two main types: distillate fuel oil and residual fuel oil. Distillate fuel oil is vaporized and condensed during a distillation process; it has a definite boiling range and does not contain high boiling oils or asphaltic components. A fuel oil that contains any amount of the residue from crude distillation hydrocracking is a residual fuel oil. However, the terms distillate fuel oil and residual fuel oil are losing their significance because fuel oils are made for specific uses and can be either distillates, residuals, or mixtures of the two. The terms domestic fuel oil, diesel fuel oil, and heavy fuel oil are more indicative of the uses of fuel oil.

Domestic fuel oils are those used primarily in the home and include kerosene, stove oil, and furnace fuel oil. Diesel fuel oils are also distillate fuel oils, but residual oils have been successfully used to power marine diesel engines, and mixtures of distillates and residuals have been used on locomotive diesels. Heavy fuel oils include a variety of oils, ranging from distillates to residual oils, that must be heated to 260°C or higher before they can be used. In general, heavy fuel oil consists of residual oil blended with distillate to suit specific needs. Heavy fuel oil includes various industrial oils and, when used to fuel ships, is called bunker oil.

Stove oil is a straight-run (distilled) fraction from crude oil whereas other fuel oils are usually blends of two or more fractions. The straight-run fractions available for blending into fuel oils are heavy naphtha, light and heavy gas oils, and residua. Cracked fractions such as light and heavy gas oils from catalytic cracking, cracking coal tar, and fractionator bottoms from catalytic cracking may also be used as blends to meet the specifications of different fuel oils.

Heavy fuel oil usually contains residuum that is mixed (cut back) to a specified viscosity with gas oils and fractionator bottoms. For some industrial purposes in which flames or flue gases contact the product (eg, ceramics, glass, heat treating, and open hearth furnaces), fuel oils must be blended to low sulfur specifications; low sulfur residues are preferable for these fuels.

Lubricating Oil. Lubricating oils are distinguished from other fractions of crude oil by their usually high (>400°C) boiling point as well as their high viscosity. Lubricating oil may be divided into many categories according to the types of service; however, there are two main groups: oils used in intermittent service, such as motor and aviation oils, and oils designed for continuous service, such as turbine oils.

Lubricating oil used in intermittent service must show the least possible variation in viscosity with respect to temperature and must be changed at frequent intervals to remove the foreign matter collected during service. The stability of such oil is therefore of less importance than the stability of oil used in continuous service for prolonged periods without renewal. Lubricating oil for continuous service must be extremely stable because the engines in which it is used operate at fairly constant temperature without frequent shutdown.

Wax. Petroleum waxes are of two general types: paraffin wax in distillates and microcrystalline wax in residua. The melting point of wax is not directly related to its boiling point because waxes contain hydrocarbons of different chemical structure. Nevertheless, waxes (qv) are graded according to their melting point and oil content. Paraffin wax is a solid crystalline mixture of straight-chain (normal) hydrocarbons ranging from mostly C₂₀ to C₃₀ and higher. Wax constituents are solid at ordinary temperatures (25°C) whereas petrolatum (petroleum jelly) contains both solid and liquid hydrocarbons.

Wax production by wax sweating was originally used in Scotland to separate wax fractions by employing various melting points from the wax obtained from shale oils. Wax sweating is still used to some extent but is being replaced by the more convenient wax recrystallization process. In wax sweating, a cake of slack wax, also known as crude or raw wax, is slowly warmed to a temperature at which the oil in the wax and the lower melting waxes become fluid and drip (or sweat) from the bottom of the cake, leaving a residue of higher melting wax.

Insofar as they are used to purify other products, several processes used in the refinery fall under the classification of dewaxing processes; however, such processes must also be classified as wax production processes (2). Most commercial dewaxing processes utilize solvent dilution, chilling to crystallize the wax, and filtration (28). The MEK process (methyl ethyl ketone–toluene solvent) is widely used. Wax crystals are formed by chilling through the walls of scraped surface chillers, and wax is separated from the resultant wax–oil–solvent slurry by using fully enclosed rotary vacuum filters.

Solvents used for dewaxing are naphtha, propane, sulfur dioxide, acetone–benzene, trichloroethylene, ethylenedichloride–benzene (Barisol), methyl ethyl ketone–benzene (benzol), methyl *n*-butyl ketone, and methyl *n*-propyl ketone. Other solvents in commercial use for dewaxing include *N*-methylpyrrolidinone, MEK–MIBK (methyl isobutyl ketone), dichloroethane–methylene dichloride, and propylene–acetone.

Solvent dewaxing can be applied to light, intermediate, and heavy lubricating oil distillates, but each distillate produces a different kind of wax, and each of these waxes is actually a mixture of a number of waxes. For example, the wax obtained from light paraffin distillate consists of a series of paraffin waxes that have melting points in the range of 30–70°C and are characterized by a tendency to harden into large crystals. However, heavy paraffin distillate yields a wax composed of a series of waxes that have melting points in the range of 60–90°C and that harden into small crystals from which they derive the name microcrystalline waxes or microwaxes.

On the other hand, intermediate paraffin distillates contain paraffin waxes and waxes intermediate in properties between paraffin and microwaxes. Thus, the solvent dewaxing process produces three different slack waxes depending on whether light, intermediate, or heavy paraffin distillate is processed. The slack wax from heavy paraffin distillate may be sold as dark raw wax, the wax from intermediate paraffin distillate as pale raw wax. The latter is treated with lye and clay to remove odor and improve color.

In the propane process, part of the propane diluent is allowed to evaporate by reducing pressure so as to chill the slurry to the desired filtration temperature, and rotary pressure filters are employed. Complex dewaxing requires no refrigeration, but depends on the formation of a solid urea–*n*-paraffin complex which is separated by filtration and then decomposed. This process is used to make low viscosity lubricants which must remain fluid at low temperatures (refrigeration, transformer, and hydraulic oils) (28).

Another method of separating petrolatum from residua is by centrifuge dewaxing. In this process, the reduced crude oil is dissolved in naphtha and chilled to –18°C or lower, which causes the wax to separate. The mixture is then fed to a battery of centrifuges where the wax is separated from the liquid. However, the centrifuge method has been largely displaced by solvent dewaxing methods as well as more modern methods of wax removal. Similar use is

anticipated for catalytic dewaxing processes based on selective hydrocracking of the normal paraffins; such processes use a molecular sieve-based catalyst in which the active hydrocracking sites are accessible only to the paraffin molecules.

Catalytic dewaxing (32) is a hydrocracking process operated at elevated temperatures (280–400°C) and pressures, 2,070–10,350 kPa (300–1500 psi). However, the conditions for a specific dewaxing operation depend on the nature of the feedstock and the product pour point required. The catalyst employed for the process is a mordenite-type catalyst that has the correct pore structure to be selective for normal paraffin cracking. Platinum on the catalyst serves to hydrogenate the reactive intermediates so that further paraffin degradation is limited to the initial thermal reactions.

Another catalytic dewaxing process also involves selective cracking of normal paraffins and those paraffins that may have minor branching in the chain. In the process (Fig. 8), the catalyst can be reactivated to fresh activity by relatively mild nonoxidative treatment. The time allowed between reactivations is a function of the feedstock; after numerous reactivations it is possible that there will be coke buildup on the catalyst.

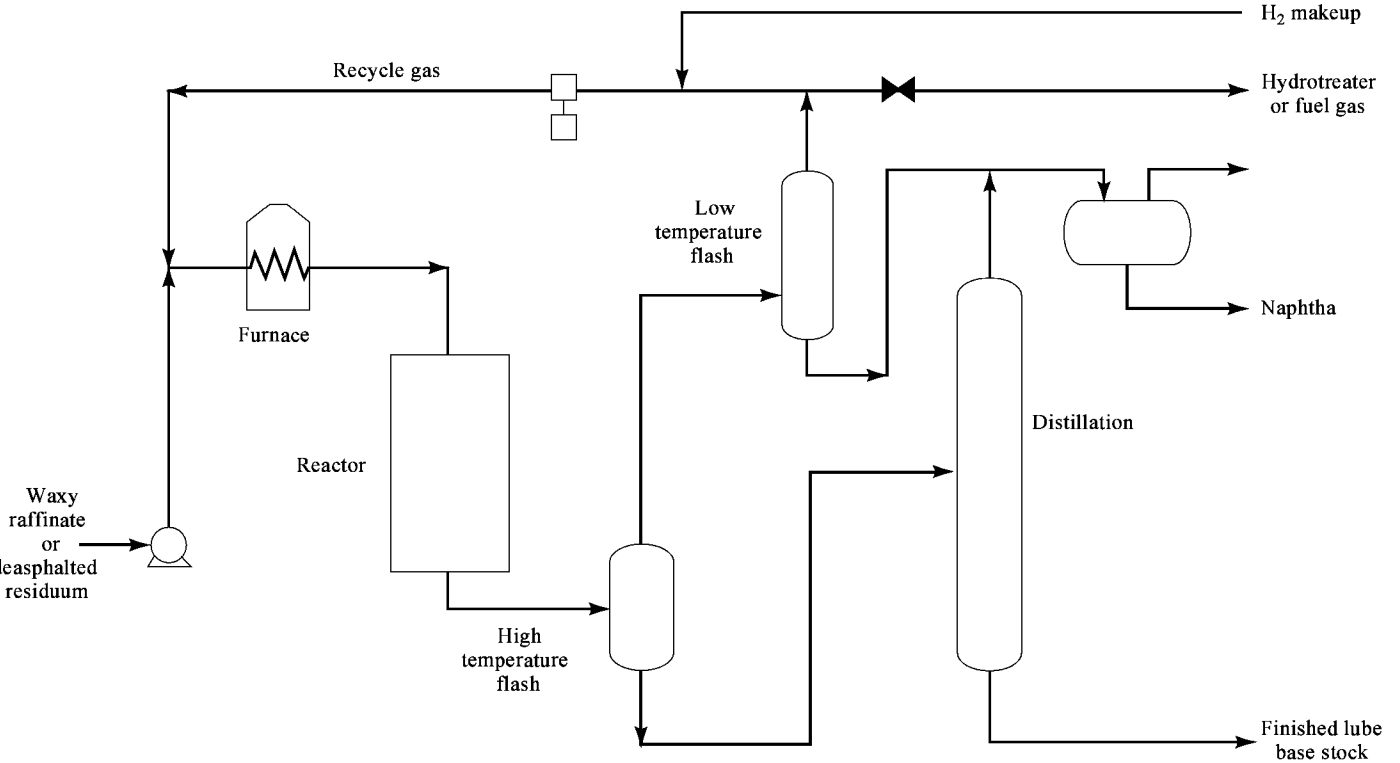


Fig. 8. Catalytic dewaxing.

A catalytic dewaxing process can be used to dewax a variety of lubricating base stocks; as such, it has the potential to replace solvent dewaxing, or even be used in combination with solvent dewaxing (Fig. 9), as a means of relieving the bottlenecks which can, and often do, occur in solvent dewaxing facilities.

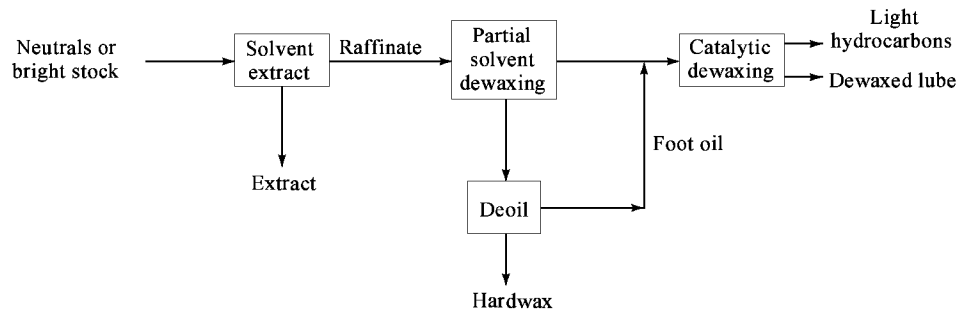


Fig. 9. Catalytic dewaxing used in conjunction with solvent dewaxing.

Asphalt. This is a distillation residuum that can also be produced by propane deasphalting (Fig. 7) (33) and thereafter modified to meet specifications. For example, asphalt (qv) can be made softer by blending hard asphalt with the extract obtained in the solvent treatment of lubricating oils. On the other hand, soft asphalts can be converted into harder asphalts by oxidation (air blowing).

Road oils are liquid asphalt materials intended for easy application to earth roads. They provide a strong base or a hard surface and maintain a satisfactory passage for light traffic. Liquid road oils, cutbacks, and emulsions are of recent date, but the use of asphaltic solids for paving goes back to the European practices of the early 1800s.

Cutback asphalts are mixtures in which hard asphalt has been diluted with a lighter oil to permit application as a liquid without drastic heating. They are classified as rapid, medium, and slow curing, depending on the volatility of the diluent, which governs the rates of evaporation and consequent hardening.

Asphalt can be emulsified with water to permit application without heating. Such emulsions are normally of the oil-in-water type. They reverse or break on application to a stone or earth surface, so that the oil clings to the stone and the water disappears. In addition to their usefulness in road and soil stabilizations, they are useful for paper impregnation and waterproofing. The emulsions are chiefly either the soap or alkaline type, or the neutral or clay type. The former breaks readily on contact, but the latter is more stable and probably loses water mainly by evaporation. Good emulsions must be stable during storage or freezing, suitably fluid, and amenable to control for the speed of breaking.

Coke. This is the residue left by the destructive distillation (coking) of residua. Petroleum coke is employed for a number of purposes; its principal use is in the manufacture of carbon electrodes for aluminum refining, which requires a high purity carbon that is low in ash and free of sulfur. In addition, coke is employed in the manufacture of carbon brushes, silicon carbide abrasives, structural carbon (eg, pipes and Rashig rings), as well as calcium carbide manufacture from which acetylene is produced. Coke produced from low quality crude oil is mixed with coal and burned as a fuel. Flue gas scrubbing is required. Coke is used in fluidized-bed combustors or gasifiers for power generation.

Fig. 11. Ethylene as a source of petrochemicals.

Ethylene (qv), an important olefin, is usually made by cracking gases such as ethane, propane, butane, or a mixture of these as might exist in a refinery's off-gases. When gas feedstock is scarce or expensive, naphthas and even whole crude oil have been used in specially designed ethylene crackers. The heavier feeds also give significant quantities of higher molecular weight olefins and aromatics. Ethylene is consumed in larger amounts than any other hydrocarbon for the production of aliphatic petrochemicals, but it is by no means the only source of aliphatic petrochemicals. Propane and butane are also important aliphatic hydrocarbons (Fig. 12). Propane is usually converted to propylene by thermal cracking, although some propylene is also available from refinery gas streams. The various butylenes are more commonly obtained from refinery gas streams. Butane dehydrogenation to butylene is known, but is more complex than ethane or propane cracking, and its product distributions are not always favorable. The production of gasoline and other liquid fuels consumes large amounts of butane.

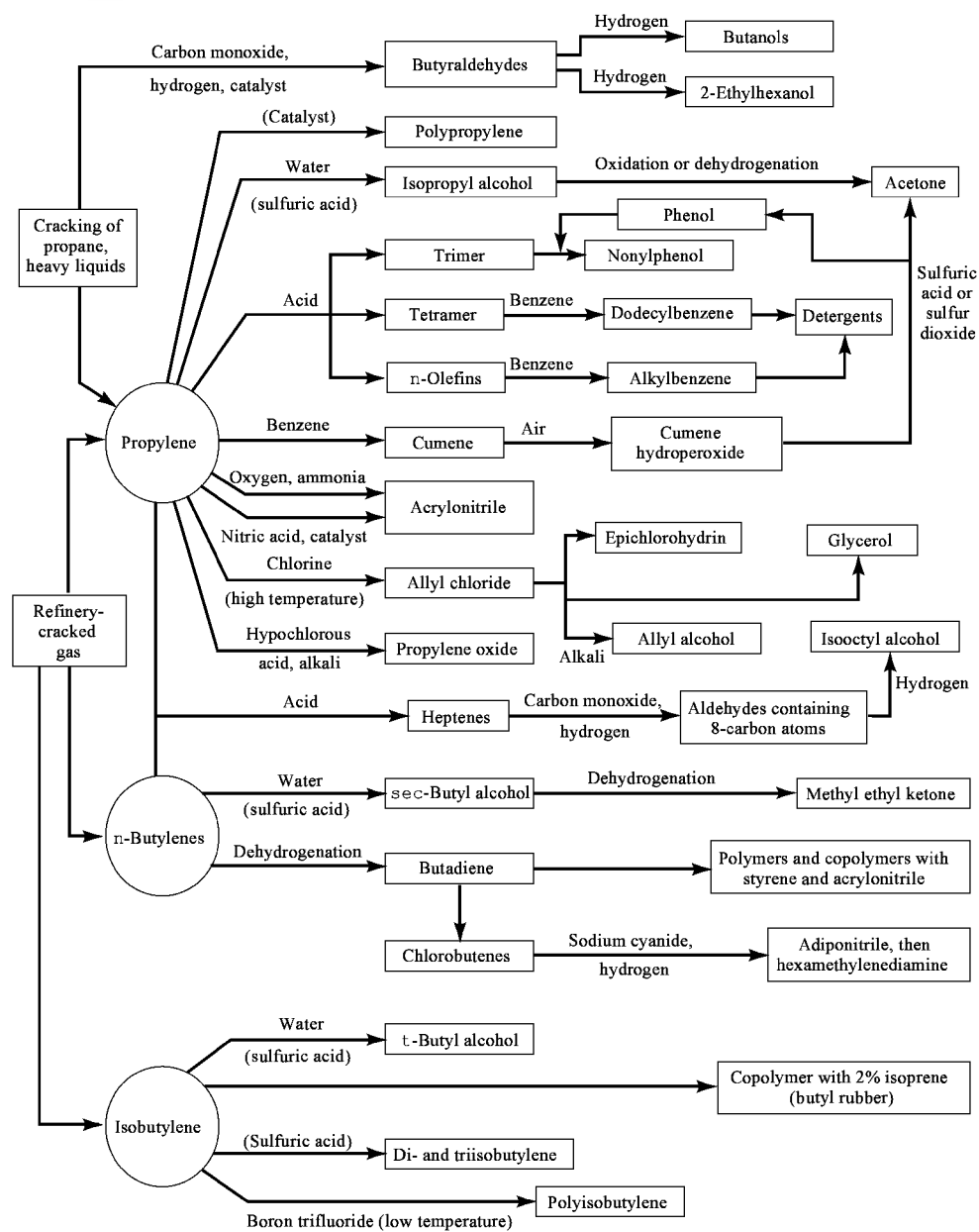


Fig. 12. Use of propylene and butylene(s) as sources of petrochemicals.

The gaseous constituents produced in a refinery give rise to a host of chemical intermediates that can be used for the manufacture of a wide variety of products (2). Synthesis gas (carbon monoxide, CO, and hydrogen, H₂) mixtures are also used to produce valuable industrial chemicals (Fig. 13).

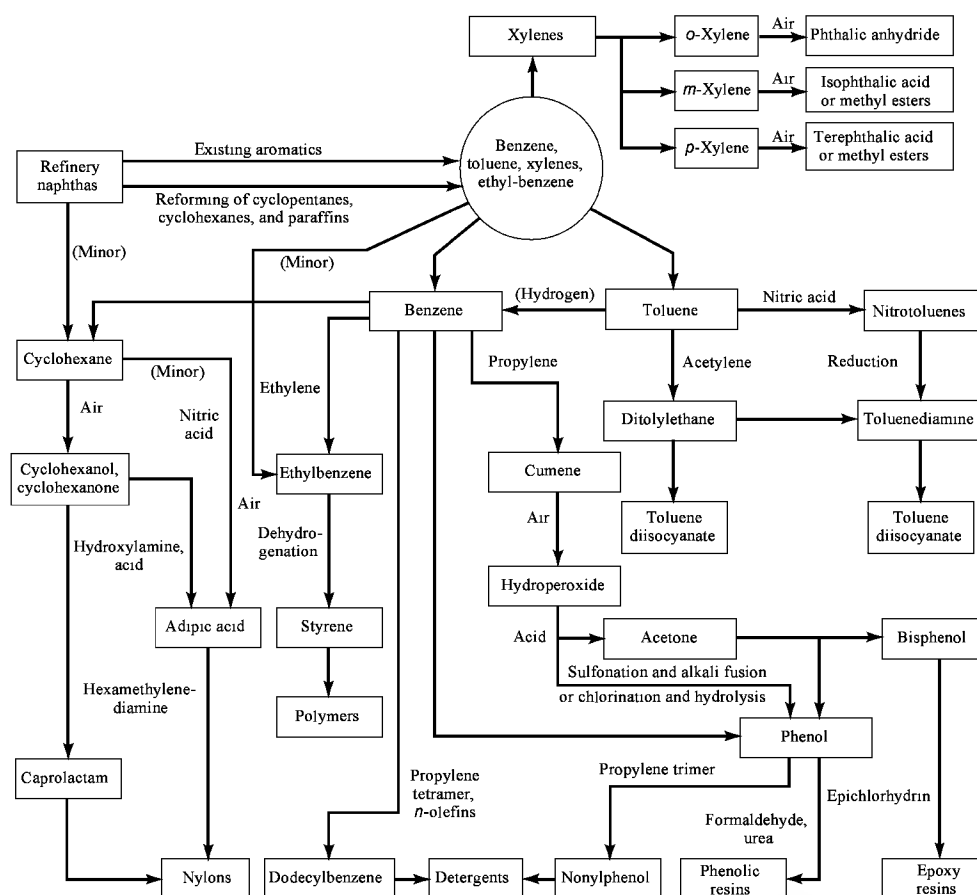


Fig. 13. Aromatic and cycloaliphatic compounds as sources of petrochemicals.

Cycloaliphatics and Aromatics. Cyclic compounds (cyclohexane and benzene) are also important sources of petrochemical products (Fig. 14). Aromatics are in high concentration in the product streams from a catalytic reformer. When aromatics are needed for petrochemical manufacture, they are extracted from the reformer's product using solvents such as glycols (eg, the Udex process) and sulfolane.

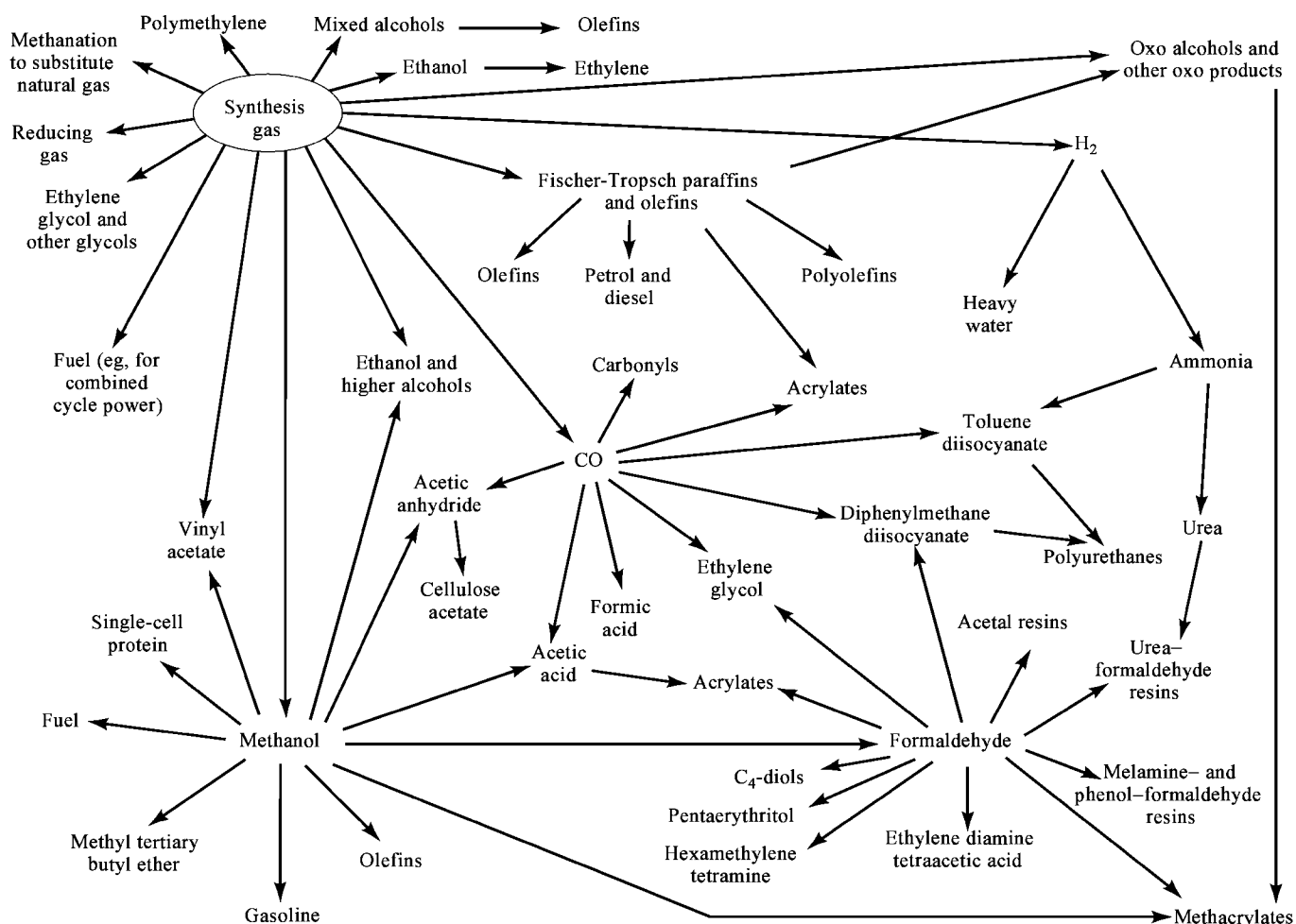


Fig. 14. Chemicals from synthesis gas.

The mixed monocyclic aromatics are called BTX as an abbreviation for benzene, toluene, and xylene (see [BTX PROCESSING](#)). The benzene and toluene are isolated by distillation, and the isomers of the xylene are separated by superfractionation, fractional crystallization, or adsorption (see [XYLENES AND ETHYLBENZENE](#)). Benzene is the starting material for styrene (qv), phenol (qv), and a number of fibers and plastics. Toluene (qv) is used to make a number of

chemicals, but most of it is blended into gasoline. Xylene use depends on the isomer: *p*-xylene goes into polyester and *o*-xylene into phthalic anhydride. Both are involved in a wide variety of consumer products.

Benzene, toluene, and xylene are made mostly from catalytic reforming of naphthas with units similar to those already discussed. As a gross mixture, these aromatics are the backbone of gasoline blending for high octane numbers. However, there are many chemicals derived from these same aromatics; thus many aromatic petrochemicals have their beginning by selective extraction from naphtha or gas–oil reformat. Benzene and cyclohexane are responsible for products such as nylon and polyester fibers, polystyrene, epoxy resins (qv), phenolic resins (qv), and polyurethanes (see FIBERS; STYRENE PLASTICS; URETHANE POLYMERS).

Inorganics. Of the inorganic chemicals, ammonia is by far the most common. Ammonia is produced by the direct reaction of hydrogen with nitrogen; air is the source of nitrogen: $\text{N}_2 + 3 \text{H}_2 \rightarrow 2 \text{NH}_3$. Refinery gases, steam reforming of natural gas (methane) and naphtha streams, and partial oxidation of hydrocarbons or higher molecular weight refinery residual materials (residua, asphalts) are the sources of hydrogen. Ammonia (qv) is used predominantly for the production of ammonium nitrate, NH_4NO_3 , as well as other ammonium salts and urea (qv), H_2NCONH_2 , which are primary constituents of fertilizers.

Carbon black, also classed as an inorganic petrochemical, is made predominantly by the partial combustion of carbonaceous (organic) material in a limited supply of air. Carbonaceous sources vary from methane to aromatic petroleum oils to coal tar by-products. Carbon black is used primarily for the production of synthetic rubber (see CARBON, CARBON BLACK).

Sulfur, another inorganic petrochemical, is obtained by the oxidation of hydrogen sulfide: $2 \text{H}_2\text{S} + \text{O}_2 \rightarrow 2 \text{H}_2\text{O} + 2 \text{S}$. Hydrogen sulfide is a constituent of natural gas and also of the majority of refinery gas streams, especially those off-gases from hydrodesulfurization processes. A majority of the sulfur is converted to sulfuric acid for the manufacture of fertilizers and other chemicals. Other uses for sulfur include the production of carbon disulfide, refined sulfur, and pulp and paper industry chemicals.

BIBLIOGRAPHY

"Petroleum Refinery Processes" in *ECT* 1st ed., Vol. 10, pp. 109–161, by W. L. Nelson and A. P. Buthod, University of Tulsa; in *ECT* 2nd ed., Vol. 15, pp. 1–77, by W. L. Nelson, University of Tulsa; "Petroleum (Refinery Processes, Survey)" in *ECT* 3rd ed., Vol. 17, pp. 183–256, by C. E. Jahnig, Consultant.

1. M. R. Gray, *Upgrading Petroleum Residues and Heavy Oils*, Marcel Dekker, Inc., New York, 1994.
2. J. G. Speight, *The Chemistry and Technology of Petroleum*, 2nd ed., Marcel Dekker, Inc., New York, 1991.
3. J. G. Speight, *The Chemistry and Technology of Coal*, 2nd ed., Marcel Dekker, Inc., New York, 1994.
4. D. R. Burris and J. D. McKinney, in J. J. McKetta, ed., *Petroleum Processing Handbook*, Marcel Dekker, Inc., New York, 1992, p. 666.
5. G. W. Mushrush and J. G. Speight, *Petroleum Products: Instability and Incompatibility*, Taylor & Francis, Washington, D.C., 1995.
6. J. G. Speight, *The Desulfurization of Heavy Oils and Residua*, Marcel Dekker, Inc., New York, 1981.
7. M. S. Matar, in H. K. Abdel-Aal, B. A. Bakr, and M. A. Al-Sahlawi, eds., *Petroleum Economics and Engineering*, 2nd ed., Marcel Dekker, Inc., New York, 1992, p. 33.
8. J. R. Salazar, in R. A. Meyers, ed., *Handbook of Petroleum Refining Processes*, McGraw-Hill Book Co., Inc., New York, 1986, pp. 7–68.
9. J. G. Gary and G. E. Handwerk, *Petroleum Refining: Technology and Economics*, Marcel Dekker, Inc., New York, 1984.
10. J. R. Salazar, in Ref. 8, pp. 7–69.
11. H. M. Feintuch, J. A. Bonilla, and R. L. Godino, in Ref. 8, pp. 7–3.
12. J. D. McKinney, in Ref. 4, p. 245.
13. D. E. Blaser, in Ref. 4, p. 253.
14. D. G. Tajbl, in Ref. 8, pp. 2–9.
15. E. C. Luckenbach and co-workers, in Ref. 4, p. 349.
16. S. A. Bradley, M. J. Gattuso, and R. J. Bertolacini, eds., *Characterization and Catalyst Development: An Interactive Approach*, American Chemical Society, Washington, D.C., 1989.
17. R. T. K. Baker and L. L. Murrell, eds., *Novel Materials in Heterogenous Catalysis*, American Chemical Society, Washington, D.C., 1990.
18. M. L. Occelli, ed., *Fluid Catalytic Cracking II: Concepts in Catalyst Design*, American Chemical Society, Washington, D.C., 1991.
19. G. E. Weismantel, in Ref. 4, p. 592.
20. D. G. Tajbl, in Ref. 8, pp. 2–33.
21. C. N. Cabrera, in Ref. 8, pp. 6–17.
22. J. A. Weiszmann, in Ref. 8, pp. 3–3.
23. P. Greenough and J. R. K. Rolfe, in Ref. 8, pp. 7–69.
24. A. Douaud, *Tomorrow's Engines and Fuels*. Report No. 94-2. Institut Francais du Petrole, Paris, 1994.
25. R. Ragsdale, *Oil Gas J.*, 51 (Mar. 21, 1994).
26. B. R. Shah, in Ref. 8, pp. 1-1.
27. T. Hutson, Jr. and W. C. McCarthy, in Ref. 8, pp. 1–23.
28. G. G. Scholten, in Ref. 4, p. 583.
29. J. G. Speight, *Gas Processing: Environmental Aspects and Methods*, Butterworth Heinemann, Oxford, U.K., 1993.
30. H. L. Hoffman, in Ref. 4, p. 2.
31. W. J. Piel, *Oil Gas J.* 40 (Dec. 4, 1989).
32. J. D. Hargrove and J. D. McKinney, in Ref. 4, p. 558.
33. C. P.-C. Chang, J. R. Murphy, and J. D. McKinney, in Ref. 4, p. 527.

James Speight
Western Research Institute

PETROLEUM RESOURCES

Petroleum resources are distributed widely in the earth's crust as gases, liquids, and solids. The products derived from these naturally occurring resources are used principally as energy sources, although substantial volumes serve as feedstocks in the chemical, plastics, and other industries (see [FEEDSTOCKS](#)). Petroleum resources are found as natural gas, as a variety of liquids that are usually classified as normal or heavy crude oils, and as semisolid and solid substances such as asphalt (qv), tar, pitch, gilsonite, and many others. The petroleum resources considered here are those liquid crude oils that can be produced through a conventional wellbore by current primary, secondary, or tertiary (enhanced recovery) production techniques and those unconventional crude oils that may be captured and converted into conventional sources of crude petroleum by advancing production technologies.

No method has been devised to estimate with complete accuracy the amount of crude petroleum that ultimately will be produced from the world's conventional oil and gas fields. Degrees of uncertainty, therefore, should be attached to all such estimates. These uncertainties can be expressed in several ways, the most important of which is achieved by dividing a resource into various categories. Several petroleum resources classifications have been proposed, and a comprehensive discussion of them (1), as well as the definition used in the assessment of the undiscovered resources of the United States (2), have been provided. Seven commonly used categories of resources are given here.

Resources represent the total amount (including reserves) of petroleum that exists in a form and amount such that economic extraction is currently or potentially feasible.

Reserves constitute the petroleum that has been discovered and can be produced at the prices and with the technology that exist when the estimate is made.

Proved reserves are estimates of petroleum reserves contained primarily in the drilled portion of fields.

Indicated reserves constitute known petroleum that is currently producible but cannot be estimated accurately enough to qualify as proved.

Inferred reserves are producible, but the assumption of their presence is based on limited physical evidence and considerable geologic extrapolation.

This places them on the borderline of being considered undiscovered, and the accuracy of the estimate is very poor.

Subeconomic resources constitute the petroleum in the ground that cannot be produced at present prices and technology but may become producible at some future date at higher prices or by improved technology.

Undiscovered resources are estimated totally by geological reasoning; no evidence through drilling is available.

To various degrees, the conventional petroleum resources in many parts of the world have been classified according to such a system. In certain regions, only estimates of proved reserves are made routinely, whereas in the United States, Canada, and certain other regions, estimates are made of volumes of petroleum in each of these categories. In the United States and Canada, estimates for several of these categories (in particular, proved reserves) are made each year by governmental agencies (3). For other categories, eg, undiscovered U.S. resources, many estimates have been prepared by various agencies, committees, panels, and companies (4–6). In a similar manner, estimates for several resource categories have been made for most other countries and for the world as a whole (7–9).

World Reserves

Most of the large volume of crude petroleum consumed in the world is extracted from only a small fraction of the total number of oil fields discovered. The concentration of crude petroleum in a few large fields is a consequence of the interaction of the geologic processes that create and trap petroleum. Even though commercial quantities of petroleum have been discovered in many localities around the world, there are enormous volume differences in fields present in a single region and in the total volume of petroleum present in different regions.

By far the largest known concentrations of conventional petroleum reserves are in the Middle East, particularly in Saudi Arabia, the United Arab Emirates, and Kuwait (Table 1). The largest concentration of reserves is in the Burgan field ($10.2 \times 10^9 \text{ m}^3$ ($64.2 \times 10^9 \text{ bbl}$)) in Kuwait (10), which contains about 68% of that country's reserves. The second largest concentration of reserves is in the Ghawar field ($7.4 \times 10^9 \text{ m}^3$ ($46.5 \times 10^9 \text{ bbl}$)) in Saudi Arabia (10), which is about 18% of that country's reserves. In some regions, a large portion of the reserves may not be contained in the largest field. However, the largest field usually contains more than 10% of the total reserves of a region. More than 20,000 petroleum fields have been discovered worldwide, and more than half of the world's proved reserves of $160.1 \times 10^9 \text{ m}^3$ ($1006.8 \times 10^9 \text{ bbl}$) of petroleum are contained in only the 51 largest fields (10).

Table 1. World Reserves and Production of Petroleum^{a,b}

Country	Reserves ^c		Production, 10^9 m^3 ^d	
	Volume, 10^9 m^3 ^d	Percentage of world	1978	1992
United States	5,103	3.2	611	514
Canada	1,208	0.8	90	120
Mexico	8,156	5.1	70	183
Total North America	14,467	9.0	771	817
Argentina	254	0.2	26	31
Brazil	477	0.3	9	40
Venezuela	9,952	6.2	126	145
others	843	0.5	45	64
Total South America	11,526	7.2	206	280
Total Western Hemisphere	25,993	16.2	977	1,097
Norway	1,399	0.9	21	126
United Kingdom	652	0.4	63	113
others	461	0.3	19	34
Total Western Europe	2,512	1.6	103	273
Iran	14,769	9.2	302	200
Iraq	15,898	9.9	153	28
Kuwait	14,944	9.3	108	53
neutral zone ^e	795	0.5	27	19
Saudi Arabia	40,986	25.6	495	507
United Arab Emirates	16,184	10.1	106	173

others	1,638	1.0	62	82
<i>Total Middle East</i>	<i>105,214</i>	<i>65.7</i>	<i>1,253</i>	<i>1,062</i>
Algeria	1,463	0.9	71	77
Libya	3,625	2.3	116	88
Nigeria	2,846	1.8	111	107
others	1,908	1.2	58	129
<i>Total Africa</i>	<i>9,842</i>	<i>6.1</i>	<i>356</i>	<i>401</i>
Australia New Zealand	318	0.2	26	34
China	3,816	2.4	116	165
Brunei Malaysia	795	0.5	26	49
former Soviet Union	9,062	5.7	651	527
India	954	0.6	15	34
Indonesia	922	0.6	95	89
other	636	0.4	25	35
<i>Total Asia</i>	<i>16,503</i>	<i>10.3</i>	<i>954</i>	<i>933</i>
<i>Total Eastern Hemisphere</i>	<i>134,071</i>	<i>83.8</i>	<i>2,666</i>	<i>2,669</i>
<i>Total World</i>	<i>160,064</i>	<i>100.0</i>	<i>3,643</i>	<i>3,766</i>

- ^a Ref. 7.
- ^b Includes crude oil, shale oil, oil sands, and where known, natural gas liquids.
- ^c January 1992.
- ^d To convert m³ to bbl, multiply by 6.29.
- ^e Neutral zone is produced jointly by Saudi Arabia and Kuwait.

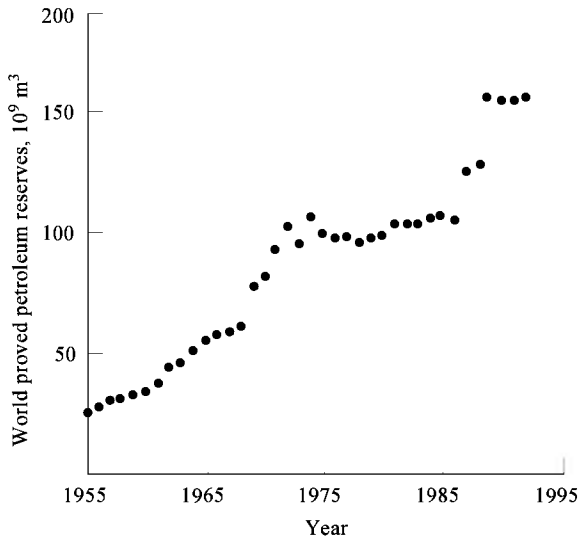


Fig. 1. Growth of world's reserves of conventional petroleum. To convert m³ to bbl, multiply by 6.29.

The world's reserves of conventional petroleum have increased from $91.7 \times 10^9 \text{ m}^3$ ($577 \times 10^9 \text{ bbl}$) in 1978 to $160.1 \times 10^9 \text{ m}^3$ ($1006.8 \times 10^9 \text{ bbl}$) in 1991 (Fig. 1). This growth of $68.4 \times 10^9 \text{ m}^3$ ($430.2 \times 10^9 \text{ bbl}$) in 13 years did not take place uniformly from region to region nor from year to year. Most of this increase took place during two years, 1987 and 1989 (Fig. 1). These increases are the result of recording of additional reserves in known fields as well as some new field discoveries, principally in the Middle East.

The U.S. proved reserves of crude petroleum and natural gas liquids (NGL) together are $5.1 \times 10^9 \text{ m}^3$ ($32.1 \times 10^9 \text{ bbl}$) and constituted 3.2% of the world's proved reserves in 1991. The U.S. position in proved reserves has fallen since 1978, when it reported $5.4 \times 10^9 \text{ m}^3$ ($34 \times 10^9 \text{ bbl}$) and constituted 6% of the world's proved reserves. Canada's proved reserves declined slightly between 1978 and 1991, whereas Mexico reported a large increase in crude petroleum reserves development over the same period, from $4.5 \times 10^9 \text{ m}^3$ ($28.3 \times 10^9 \text{ bbl}$) to $8.2 \times 10^9 \text{ m}^3$ ($51.6 \times 10^9 \text{ bbl}$), thereby surpassing the United States and becoming the country with the largest proved reserves in North America (see Table 1).

In South America, Venezuela continues to dominate in the proved reserve and the production categories. Since 1978, reserves of crude petroleum in Venezuela have increased from $2.9 \times 10^9 \text{ m}^3$ ($18.2 \times 10^9 \text{ bbl}$) to $10.0 \times 10^9 \text{ m}^3$ ($62.6 \times 10^9 \text{ bbl}$), nearly doubling its share of the world's proved reserves from 3.2 to 6.2%. Venezuela has extra large deposits of heavy oils in the East Venezuela Basin, the further development of which may expand its position in proved petroleum reserves.

The 1991 petroleum resources of Western Europe were almost identical to what they were in 1978 ($2.5 \times 10^9 \text{ m}^3$ ($16.1 \times 10^9 \text{ bbl}$) vs $2.6 \times 10^9 \text{ m}^3$ ($15.8 \times 10^9 \text{ bbl}$)). During that period, the net decline in reserves in the United Kingdom from $1.6 \times 10^9 \text{ m}^3$ ($10.2 \times 10^9 \text{ bbl}$) to $0.7 \times 10^9 \text{ m}^3$ ($4.1 \times 10^9 \text{ bbl}$) was offset by the net increase in reserves in Norway from $0.7 \times 10^9 \text{ m}^3$ ($4.1 \times 10^9 \text{ bbl}$) to $1.4 \times 10^9 \text{ m}^3$ ($8.8 \times 10^9 \text{ bbl}$). The reserves of the rest of Western Europe are insignificant (Table 2).

At the end of 1991, the reserves of crude petroleum in Africa were $9.8 \times 10^9 \text{ m}^3$ ($61.9 \times 10^9 \text{ bbl}$), or only slightly higher than those in 1978, when they were $8.9 \times 10^9 \text{ m}^3$ ($56.3 \times 10^9 \text{ bbl}$). Algeria, Libya, and Nigeria account for over 80% of these reserves and over 65% of the production from Africa.

The reserves of crude petroleum in Asia stood at $16.5 \times 10^9 \text{ m}^3$ ($103.8 \times 10^9 \text{ bbl}$) in 1991. This is an increase of 10% since 1978; most of this increase was accounted for by China, India, and Brunei Malaysia. Levels of proved reserves fell during this period in Australia New Zealand, Indonesia, and several other Asian countries. In the countries that formerly composed the Soviet Union, reserves decreased slightly (2.5%) between 1978 and 1991, whereas annual production decreased 19%. For many years, the Soviet Union had been the leading producer of crude petroleum in the world, a position it still held in 1991, when it produced $527 \times 10^6 \text{ m}^3$ ($3.3 \times 10^9 \text{ bbl}$). This level is only slightly higher than production levels in the United States ($514 \times 10^6 \text{ m}^3$ ($3.2 \times 10^9 \text{ bbl}$)) and Saudi Arabia ($507 \times 10^6 \text{ m}^3$ ($3.2 \times 10^9 \text{ bbl}$)).

The proved reserves and levels of production for Japan, Myanmar (formerly Burma), Pakistan, Taiwan, and Thailand are insignificant by world standards. In 1979, the Philippines established the first commercial production in the small offshore South Nido field. This success came after more than 75 years of wildcat drilling in the Philippines. After several additional discoveries, production rose to $0.3 \times 10^6 \text{ m}^3$ ($1.7 \times 10^6 \text{ bbl}$) in 1991.

Table 2. Estimates of Proved Petroleum Reserves in Parts of Western Europe

Country	Volume, 10 ³ m ³ ^a
Denmark	116.1
Italy	110.1
Germany	43.2
former Yugoslavia	29.7
Spain	2.9
Austria, Greece, the Netherlands, and France	69.2
<i>Total</i>	<i>371.2</i>

^a To convert m³ to bbl, multiply by 6.29.

U.S. Reserves

Between 1978 and 1991, U.S. proved reserves of crude petroleum decreased by 21.3% from 5.0 × 10⁹ m³ (31.4 × 10⁹ bbl) to 3.9 × 10⁹ m³ (24.7 × 10⁹ bbl), as listed in Table 3. During this same period, NGL reserves increased by 33% from 0.9 × 10⁹ m³ (5.9 × 10⁹ bbl) to 1.2 × 10⁹ m³ (7.5 × 10⁹ bbl). The data in Table 3 show that despite small net additions in several U.S. states, eg, Colorado and New Mexico, the conventional crude petroleum reserves of the United States were depleted rapidly between 1978 and 1991. Even with this decline in proved reserves, the United States was the second largest producer of crude petroleum in the world in 1992 after the former Soviet Union (see Table 1). Although much crude petroleum in the United States in recent years has been credited to the proved inventory through the extension and revision development processes, many of the newer discoveries of conventional hydrocarbon have been natural gas (see GAS, NATURAL).

Table 3. Estimates of Proved Reserves and Production of Crude Petroleum in the United States, 10⁶ m^{3a,b}

States	Proved reserves		1978–1991	
	End 1978 ^c	End 1991 ^c	Net additions to reserves	Cumulative production ^d
Alabama	11.7	6.8	-4.9	41.9
Alaska	1491.9	967.1	-524.8	1329.7
Arkansas	17.6	11.1	-6.5	31.4
California	790.8	795.2	+4.4	844.5
Colorado	30.8	52.3	+21.5	63.4
Florida	26.7	5.9	-20.8	39.8
Illinois	25.1	20.3	-4.8	48.4
Indiana	4.6	2.5	-2.1	8.6
Kansas	48.2	47.7	-0.5	138.3
Kentucky	6.4	4.9	-1.5	10.0
Louisiana	548.2	390.1	-158.1	963.1
Michigan	35.0	18.9	-16.1	56.8
Mississippi	39.7	30.8	-8.9	67.3
Montana	25.1	32.0	+6.9	55.8
Nebraska	4.8	4.1	-0.7	87.5
New Mexico	92.1	114.6	+22.5	154.4
North Dakota	25.8	36.9	+11.1	87.5
Ohio	11.0	10.5	-0.5	23.4
Oklahoma	155.6	111.3	-44.3	295.7
Pennsylvania	4.3	2.4	-1.9	5.1
Texas	1416.7	1090.1	-326.6	1759.0
Utah	29.9	37.0	+7.1	63.7
West Virginia	4.8	4.1	-0.7	6.2
Wyoming	134.3	120.3	-14.0	246.4
miscellaneous ^e	3.8	6.7	+2.9	10.9
<i>Total</i>	<i>4984.9</i>	<i>3923.6</i>	<i>-1061.3</i>	<i>6438.8</i>

^a Ref. 3.
^b To convert m³ to bbl, multiply by 6.29.
^c December 31.
^d Includes lease condensates.
^e Includes Arizona, Maryland, Missouri, Nevada, New York, South Dakota, Tennessee, and Virginia.

Ultimate Petroleum Resources of the World

Since the late 1960s, the ultimate amount of crude petroleum in the world that is producible through conventional production techniques has been estimated to be about 350 × 10⁹ m³ (2.2 × 10¹² bbl) (9,11–13). By the end of 1991, cumulative world production was 103.8 × 10⁹ m³ (652.9 × 10⁹ bbl) (14), and world proved reserves were estimated to be 160.1 × 10⁹ m³ (1006.8 × 10⁹ bbl) (see Table 1). Thus, by the end of 1991, 263.9 × 10⁹ m³ (1659.7 × 10⁹ bbl) of crude petroleum had been discovered, which is more than 75% of the estimated 350 × 10⁹ m³ (2200.0 × 10⁹ bbl) of conventional crude petroleum estimated to be ultimately recoverable.

World Petroleum Supply and Consumption

As shown in Table 4, the 1992 world consumption of petroleum was nearly 10.4 × 10⁶ m³ d (65.4 × 10⁶ bbl d) (8), which is slightly higher, at 3.6%, than in 1978. In most of the regions shown in Table 4, consumption and production levels are not in balance. The one exception is the group of non-OECD European countries, ie, the former Soviet Union, the former Czechoslovakia, Hungary, Poland, and other former Eastern Bloc countries, where production was nearly in balance with consumption at about 1% above consumption in 1992.

Table 4. World Petroleum Statistics, 1992, 10³ m³/d^{a,b}

Area	Consumption	Production ^c	Refinery capacity ^d
North America	2,847	1,735	2,732
Latin America	829	1,269	1,211
<i>Total Western Hemisphere</i>	<i>3,676</i>	<i>3,004</i>	<i>3,943</i>
OECD Europe	2,169	747	2,226
non-OECD Europe ^e	1,250	1,485	2,044
Middle East	560	2,909	781
Africa	322	1,100	465
Asia Australia	2,419	1,076	2,308
<i>Total Eastern Hemisphere</i>	<i>6,720</i>	<i>7,317</i>	<i>7,824</i>
<i>Total World</i>	<i>10,396^f</i>	<i>10,321^f</i>	<i>11,767</i>

^a Ref. 7.

^b To convert m³ to bbl, multiply by 6.29.

^c Includes U.S. natural gas liquids.

^d On December 31, 1992.

^e Includes the former Soviet Union, the former Czechoslovakia, Hungary, Poland, and other former Eastern Bloc countries.

^f Differences between production and consumption are accounted for by stock change and unknown military liftings.

Consumption in North America in 1992 ($2.8 \times 10^6 \text{ m}^3 \text{ d}$ ($17.9 \times 10^6 \text{ bbl d}$)), although lower than that of 1978 ($3.2 \times 10^6 \text{ m}^3 \text{ d}$ ($20.1 \times 10^6 \text{ bbl d}$)), nonetheless exceeded production in the region by 64^o o. In Latin America, consumption rose to $829.0 \times 10^3 \text{ m}^3 \text{ d}$ ($5.2 \times 10^6 \text{ bbl d}$) in 1992, up from $666.0 \times 10^3 \text{ m}^3 \text{ d}$ ($4.2 \times 10^6 \text{ bbl d}$) in 1978, or an increase of 24^o o; production in this region increased by 61^o o from $789.0 \times 10^3 \text{ m}^3 \text{ d}$ ($5.0 \times 10^6 \text{ bbl d}$) in 1978 to $1269.0 \times 10^3 \text{ m}^3 \text{ d}$ ($8.0 \times 10^6 \text{ bbl d}$) in 1992. Much of this increase in production is from fields discovered in the 1980s in Brazil and Colombia.

Historically, the world's petroleum production pattern can be related to geologic, economic, and political factors. In the past, many countries have had large excesses in production capacity, whereas in the 1990s, only countries in the Middle East, such as Saudi Arabia, Kuwait, Iraq, and the United Arab Emirates, have, in the short run, enough excess capacity to expand production of conventional crude petroleum in any significant manner. In the Middle East, production of petroleum is over five times the region's consumption (see Table 4). On a much smaller scale, Africa produces far more petroleum than it consumes (3.4 times).

In terms of consumption in the Eastern vs the Western Hemisphere, the data in Table 4 show that not only is most of the world's petroleum produced in the Eastern Hemisphere (71^o o), but it is also consumed largely in that region (65^o o), with 8.5^o o in Japan alone. The surplus production in the Eastern Hemisphere ($597 \times 10^3 \text{ m}^3 \text{ d}$ ($3.8 \times 10^6 \text{ bbl d}$)) is consumed in the Western Hemisphere, mostly in North America. The Western Hemisphere thereby produces 29^o o of the world's total production, or about 82^o o of the petroleum that it consumes.

For some time, annual production and reserves of conventional petroleum have remained nearly level or have been declining in a number of countries, eg, the United States, Canada, the United Kingdom, Algeria, Australia New Zealand, the former Soviet Union, Indonesia, and a number of other, smaller producers. Unless substantial new reserves can be discovered in these countries, the production rates will continue to decline. The rate of these declines will be determined by the physical quality of the existing reservoirs. Some anticipated decline in production levels in these countries may be prevented in the short run by more intense reservoir development, eg, in-field drilling and application of enhanced recovery technology. Reserves in several Middle Eastern countries are large enough to support substantially increased production, as are reserves in Venezuela, Colombia, Norway, and several other countries.

Perhaps the most striking feature shown in Table 4 is the large imbalance between regional production and consumption in the Middle East as compared to OECD Europe. In 1992, the Middle East produced five times more crude petroleum than it consumed, and OECD Europe consumed about three times more crude petroleum than it produced; that is, in 1992, the Middle East exported about 81^o o of its production of crude petroleum, whereas OECD Europe imported about 66^o o of the crude petroleum that it consumed. In the Asia Australia region, $2419 \times 10^3 \text{ m}^3 \text{ d}$ ($15.2 \times 10^6 \text{ bbl d}$), or 23.3^o o of the world's total, was consumed in 1992. In the United States, production of crude petroleum peaked in 1971 and has declined since then so that only 54.5^o o of the U.S. crude petroleum consumed in 1992 was produced domestically. In Latin America, production of crude petroleum stood at about 150^o o of consumption, whereas in 1978, production and consumption were about equal.

Outlook

Petroleum displaced coal (qv) as the principal source of energy in the United States by 1948 and in the world by 1965 (15). In 1992, petroleum satisfied over 40^o o of the world's energy needs, while coal filled only 28^o o of needs, barely ahead of natural gas at 23^o o. The spectacular growth in consumption of crude petroleum in the world during the middle and late twentieth century is directly attributable to the ease with which petroleum can be discovered, produced, transported, processed, and utilized (see PETROLEUM, ENHANCED OIL RECOVERY). This growth has been so rapid that as much crude petroleum ($55.5 \times 10^9 \text{ m}^3$ ($349.4 \times 10^9 \text{ bbl}$)) was taken from the ground between 1976 and 1992 as was produced during the entire previous 119-yr period (1857–1975). This rapid rate of expansion in production and consumption, coupled with the finiteness of the conventional petroleum resource base, has from time to time led some analysts to conclude that world petroleum production will peak in the near future (16,17). Other analysts who examine such data forecast impending global crisis as crude petroleum consumption declines and coal reclaims its former position as the principal source of fossil energy (18,19).

The key factor influencing the varying interpretations is that although there is an enormous volume of petroleum resources in the ground throughout the world, it is found in deposits that differ in quality and quantity from country to country. Only a small fraction of these resources are conventional petroleum resources ($160.1 \times 10^9 \text{ m}^3$ ($1006.8 \times 10^9 \text{ bbl}$)) and are in the category of proved reserves. An additional $93 \times 10^9 \text{ m}^3$ ($585 \times 10^9 \text{ bbl}$) of conventional petroleum is estimated to be undiscovered in the world. This estimate is of undiscovered petroleum resources that are economic to produce by means of normal production technology.

Although the world is not running out of petroleum, it is difficult to sum up how much is available in the short run, as well as in the longer run, in light of various possible future political and economic developments. The effect of a variety of sociopolitical forces now at work will be to reduce the world's consumption of energy, in particular of petroleum; such forces include clean-air regulations in the United States and the ever-increasing rate of taxation of petroleum use in many countries in order to raise general revenues. Countervailing forces, such as the advancement of exploration and production technology, can be counted on to expand the discovery and development of additional conventional petroleum resources in deep-water offshore regions and in hostile arctic climates. Also, advancements in technology, eg, in three-dimensional seismic surveys and horizontal drilling, will increase the inventory of proved reserves through the process of field extensions. Over the next several decades, many billions of cubic meters of conventional petroleum will be credited to the reserves inventory through this field-growth process.

Perhaps the biggest contribution that technological advancement in petroleum production will make is bringing large volumes of unconventional petroleum resources, eg, heavy oil and tar sands, into a viable economic realm by lowering the unit cost of production. Compared to the inventory of conventional petroleum reserves and undiscovered resources, the physical inventories of such unconventional petroleum resources are extremely large; for example, the Athabasca tar sands in Alberta, Canada, are estimated to contain $360 \times 10^9 \text{ m}^3$ ($2250 \times 10^9 \text{ bbl}$) of in-place petroleum (19). This volume is equivalent to the total inventory, ie, the combined cumulative production, reserves, and undiscovered resources, of world conventional crude petroleum. In

1992, however, only about 10% of total in-place petroleum resources was technically recoverable (20).

Large unconventional resources of petroleum also occur as extra heavy crude oils in the Orinoco belt, Venezuela, and in oil shale in the western United States. Petroleum resources in the unconventional category, such as tar sands, heavy crude oils, and oil shales, are located mostly in the Western Hemisphere, as opposed to the conventional resources, which are located mostly in the Middle East. Also, the in-place resources of these unconventional resources are about twice as large as the in-place resources of conventional crude petroleum. Although the recovery rates from these resources are low, improving technology may capture increasing volumes of these unconventional petroleum resources, thereby converting them into conventional petroleum resources.

The irony underlying the current perceptions of the world's petroleum resource situation is that the world is not running out of combined petroleum resources. However, the bulk of the world's conventional petroleum resources that are inexpensive to produce are concentrated in the Middle East, whereas the much larger volume of higher cost unconventional petroleum resources, such as tar sands and heavy oils, are located in Venezuela, western Canada, and the western United States. The uneven distribution of the conventional resources thus causes concern that in the short run political unrest in the Middle East could cause a catastrophic interruption of supply.

However, the general pessimism about a perceived threat to the world economy that existed at the time of the Arab oil embargoes in the 1970s has been dispelled. At the core of this concern was the idea that the well-being of modern society was totally dependent on the ever-expanding use of cheap energy, especially crude petroleum. Since that time, the world has consumed petroleum at the nearly steady rate of $3.5 \times 10^9 \text{ m}^3 \text{ yr}$ ($22 \times 10^9 \text{ bbl yr}$) even as the world economy has continued to expand. This ending of growth in consumption is the result of the price rises caused by the Arab oil embargoes in 1970s, which have prompted the world's economic agents to turn to conservation, technology, and in some situations interfuel substitution, eg, alcohol for gasoline, and coal for petroleum in electric power generation (qv). Perhaps the most surprising response to the increasing price of petroleum was the speed with which the world downsized its fleet of private automobiles, thereby lowering the demand for petroleum in that sector.

The world will never "run out" of petroleum, simply because there is so much of it in the ground in so many different forms. However, the resources of conventional crude petroleum are finite. These are the petroleum resources that are very inexpensive to produce because they flow to the wellbore either directly or by pumping after the application of standard well completion methods. There is a more or less general agreement among analysts that the size of the inventory of these resources is about $350 \times 10^9 \text{ m}^3$ ($2200 \times 10^9 \text{ bbl}$); the world is consuming these resources at about 1% yr. The primary question that faces the world is, "Has the pattern of exponential growth in consumption of petroleum that took place between the end of World War II and 1973 become a relic of the past, or could growth resume as world population continues to expand?" Analysis of the pattern of world energy consumption shows that the world consumption of crude petroleum may gradually increase even with increased efficiency in the use of energy, simply as a result of population growth. However, these developments could be dramatically altered by an increase in the price of energy (21).

Another consideration of petroleum assessment analysts is whether, and to what degree, the vast resources of unconventional petroleum in the world can be captured by advances in petroleum production technologies, thereby converting them into conventional sources of petroleum. It is a simple fact that the in-place resources of petroleum in tar sands, heavy oils, and oil shale can guarantee the future supply of petroleum for hundreds of years at the current rate of consumption, provided they can be produced at competitive costs.

BIBLIOGRAPHY

"Petroleum Resources" under "Petroleum" in *ECT* 2nd ed., Vol. 14, 856–867, by E. R. Heydinger, Marathon Oil Co.; "Resources" under "Petroleum" in *ECT* 3rd ed., Vol. 17, pp. 132–142, by L. J. Drew, U.S. Geological Survey.

1. J. J. Schanz, Jr., *1978 Oil and Gas Resources—Welcome to Uncertainty*, Resources (Special Issue), Resources For The Future, Washington, D.C., 1978, 16 pp.
2. U.S. Geological Survey—U.S. Minerals Management Service, *Estimates of Undiscovered Conventional Oil and Gas Resources in the United States—A Part of the Nation's Energy Endowment*, unnumbered report, U.S. Geological Survey and the U.S. Minerals Management Service, Washington, D.C., 1989, 44 pp.
3. Energy Information Administration, *U.S. Crude Oil, Natural Gas, and Natural Gas Liquids Reserves, 1991*, DOE EIA-0216(91), U.S. Department of Energy, Washington, D.C., 1991, 129 pp.
4. D. Gautier and co-workers, *Estimates of Undiscovered Conventional Oil and Gas Resources in the United States—A Part of the Nation's Energy Endowment*, USGS Research on Energy Resources (V. E. McKelvey Forum), Washington, D.C., Feb. 1995.
5. E. R. Schroeder, in D. D. Rice, ed., *Oil and Gas Resources in the U.S. Arctic* (by the National Petroleum Council), No. 21, American Association of Petroleum Geological Studies in Geology, 1986, pp. 111–123.
6. R. A. Rozendal, in Ref. 5, pp. 151–158.
7. C. D. Masters, in Ref. 5, pp. 179–184.
8. *BP Statistical Review of World Energy*, British Petroleum Ltd., London, June 1993.
9. C. D. Masters, D. H. Root, and E. D. Attanasi, in *Proceedings of 13th World Petroleum Congress, Buenos Aires*, topic 25, 1991, pp. 1–13.
10. Technical data, Petroconsultants SA, Geneva, Switzerland, 1992.
11. M. Grenon, *World Oil Resources—Assessment and Potential for the 21st century*, Preprint for U.S. Geol. Int. Resources Symposium, Reston, Va., Oct. 1979.
12. D. C. Ion, *Availability of World Resources*, Graham and Trotman Ltd., London, 1976.
13. H. R. Warman, *Petro. Rev.* **5**, 96–99 (1971).
14. *DeGolyer and MacNaughton Twentieth Century Petroleum Statistics*, Dallas, Tex., 1992, 126 pp.
15. H. Enzer, W. Dupree, and S. Miller, *Energy Perspectives*, U.S. Dept. of the Interior, Washington, D.C., 1975.
16. M. K. Hubbert, in *Resources and Man*, National Geological Sciences, National Research Council, Washington, D.C., 1969, Chapt. 8.
17. D. H. Root and E. D. Attanasi, *Natl. Res. Forum* **4**(2), 181 (1980).
18. W. S. Fyfe, M. A. Powell, B. R. Hart, and B. Ratanasthien, *Nonrenewable Res.* **2**(3), 189 (1993).
19. C. D. Masters, E. D. Attanasi, and D. H. Root, "World Petroleum Reserves and Analysis," *World Petroleum Congress Proceedings*, in press.
20. N. Alazard and L. Montadert, *Nonrenewable Res.*, **2**(3), 197–206 (1993).
21. J. F. Brookout, *Episodes* **12**(4), 257–262 (1989).

Lawrence J. Drew
U.S. Geological Survey

PETROLEUM PRODUCTS.

See FEEDSTOCKS,PETROCHEMICAL.

PHARMACEUTICALS

Pharmaceuticals are best viewed as drug-containing products in dosage forms. These forms are designed and manufactured to deliver safe and effective therapeutic responses each time administered within appropriate regimens and even after storage under well-documented conditions in scientifically designed packaging for designated time periods (see PACKAGING, COSMETICS AND PHARMACEUTICALS). Thus, pharmaceuticals are actual drug delivery systems (qv). Pharmacokinetic and pharmacodynamic principles that influence the delivery of a drug from the pharmaceutical product to the body, routes of drug administration, and modes technologies used for controlled drug delivery are covered elsewhere (see DRUG DELIVERY SYSTEMS; PHARMACODYNAMICS).

Various technologies are required to produce drug products. Both federal and state laws and regulations exist in the United States to control the manufacture and distribution of pharmaceuticals. The U.S. drug distribution system is multifaceted including drug usage within the community and hospitals, under long- or short-term home health care or pharmacy practice. Individual pharmaceuticals are covered elsewhere in the *Encyclopedia*. These include ANALGESICS, ANTIPYRETICS, AND ANTIINFLAMMATORY AGENTS; ANESTHETICS; ANTIBACTERIAL AGENTS, SYNTHETIC and ANTIBIOTICS; ANTIPARASITIC AGENTS; appetite-suppressing agents (see ANTI Obesity DRUGS); ANTIasthmatic AGENTS; CARDIOVASCULAR AGENTS; chemotherapeutics (see CHEMOTHERAPEUTICS, ANTICANCER); CONTRACEPTIVES; DISINFECTANTS AND ANTISEPTICS; DIURETICS; EPINEPHRINE AND NOREPINEPHRINE; EXPECTORANTS AND ANTITUSSIVES; GASTROINTESTINAL AGENTS; histamine and antihistamine agents (see HISTAMINE AND HISTAMINE ANTAGONISTS); HORMONES; HYPNOTICS, SEDATIVES, ANTICONVULSANTS, AND ANXIOLYTICS; IMMUNOTHERAPEUTIC AGENTS; INSULIN AND OTHER ANTIDIABETIC AGENTS; MEMORY ENHANCING DRUGS and ANTIAGING AGENTS; NEUROREGULATORS; controlled release dosage forms (see CONTROLLED RELEASE TECHNOLOGY, PHARMACEUTICAL); optically active dosage forms (see PHARMACEUTICALS, CHIRAL); PROstaglandins; PSYcopharmacological AGENTS; STEROIDS; THYROID AND ANTITHYROID PREPARATIONS; VETERINARY DRUGS; and VITAMINS.

Economic Aspects

The pharmaceutical industry comprises a significant economic and research component of chemical technology in the United States. Total global sales for 1993 for the industry exceeded \$95 billion (1–5). The Pharmaceutical Research and Manufacturers of America (PhRMA), formerly the Pharmaceutical Manufacturers Association, estimated 1993 sales of human-use and veterinary-use ethical (prescription) pharmaceuticals sold by its member companies to be approximately \$57 billion domestically and approximately \$27.9 billion abroad (1). The U.S. Generic Pharmaceutical Industry Association (GPIA) estimated generic U.S. drug sales for 1993 to be approximately \$10.1 billion (2). The GPIA estimates generics to comprise >43% of the U.S.-dispensed prescriptions (2). The Nonprescription Drug Manufacturers Association (NDMA) reported U.S. sales of \$13.3 billion for over-the-counter (OTC) products during 1993 and \$13.8 billion in 1994. Sales of pharmaceuticals (ethical, generic, and OTC) for 1993 to U.S. retail outlets that have component pharmacies (independent community, regular chain, and food markets), with the exception of mail-order pharmacies, have been reported as \$486 billion. Sales to U.S. hospitals were ca \$9.9 billion (5).

PhRMA is a trade association of over 100 research-based pharmaceutical companies. For membership a company must manufacture and market finished dosage-form products under its own brand names and must conduct a significant amount of research and development in the United States.

U.S. Pharmacy: Licensure and Principal Organizations

In the United States, there is no national qualifying or licensing body for pharmacists. Licensure requirements are promulgated by State boards of pharmacy that administer examinations, issue internship requirements, and oversee the practice of pharmacy. The National Association of Boards of Pharmacy serves the collective needs of the state boards. This organization has no licensure authority. However, it has developed a standardized licensure examination (NABPLEX), which as of this writing (ca 1995) is used by 48 states (see LICENSING).

Several national organizations serve the professional needs of U.S. pharmacists. These reflect the practice milieu of members, eg, independent community pharmacies, chain drug stores, and hospitals. The American Pharmaceutical Association (APhA), founded in 1852, is composed of the Academy of Pharmaceutical Research and Science, Academy of Pharmaceutical Practice and Management, and the Academy of Students of Pharmacy. Other organizations include the American Society of Health-Systems Pharmacists (ASHP), National Association of Chain Drug Stores (NACDS), and National Association of Retail Druggists (NARD).

The American College of Apothecaries represents pharmacists whose practices can best be described as emphasizing prescription and related products. Some pharmacists practice as consultants and providers to long-term care health facilities, eg, nursing homes. Both state and U.S. laws have mandated closer control of drug products in such units. The American Association of Consultant Pharmacists has been formed to serve the needs of such pharmacists.

The pharmaceutical industry is represented by several organizations: the Pharmaceutical Research and Manufacturers of America (nñe: Pharmaceutical Manufacturers of America), the Non-Prescription Drug Manufacturers Association, and the National Pharmaceutical Council. The schools and colleges of pharmacy are organized as the American Association of Colleges of Pharmacy, representing both schools and colleges, and faculty members. Scientific professionals involved in the research, development, and manufacture of pharmaceuticals can become members of the American Association of Pharmaceutical Scientists. The industry, through the American Foundation for Pharmaceutical Education, supports education via various programs, including graduate fellowships. The American Council for Pharmaceutical Education is the accrediting body for U.S. colleges and schools of pharmacy. It also accredits continuing-education providers.

Each state has a professional pharmacy organization, some of which are affiliated with the American Pharmaceutical Association. Similarly, state organizations of hospital pharmacists exist in affiliation with the ASHP. Likewise, local or county associations exist in most instances. Each national association publishes a journal as do most state organizations. The *Federal Register* reports proposed and enacted federal regulatory occurrences several times a week. Each state has a similar publication to report its legislation and regulatory developments, eg, *The Pennsylvania Bulletin*.

Drugs and Drug Products

The U.S. Food and Drug Administration (FDA) approved 22 new drugs and one biotech medicine during 1994. These new drug entities had an adjusted average review time of 19.7 months, from filing of the New Drug Application (NDA) at the FDA to time of approval. This was down from the 25.6 months for the 26 new entities approval in 1993. In the total drug development and approval process it takes approximately 12 years for an experimental

drug to go from the lab to the medicine chest (6). Only about 5 in 5000 new chemical entities that enter preclinical (lab and animal studies) testing reach human clinical testing (Phase I, II, and III) and only one of the five tested clinically is approved. On average a pharmaceutical manufacturer invests ~\$360 million to get one new drug to the consumer or patient.

In the United States, through the NDA review process, pharmaceutical companies that seek FDA approval for new drug products are assessed user fees by FDA to gain faster approval, by virtue of the U.S. Prescription Drug User Fee Act of 1992. These assessments are used to increase the new drug review staff of the FDA, which has agreed to reduce the NDA review time to 12 months by 1997 (6).

In 1962, amendments to the U.S. Federal Food, Drug and Cosmetic Act (Kefauver-Harris amendments) promulgated regulations concerning the requirements for premarketing approval by the FDA. This legislation established requirements of proof of both safety and therapeutic efficacy and strict control of human clinical testing, for example, which have extended the time and cost to market a new drug. Thus, whereas approximately 40 new drugs were marketed annually from 1948 to 1962, this number had fallen to 12 by 1966.

The increase in time to bring a new drug to the point of FDA approval, that the 1962 amendments generated, reduced the length of the effectiveness of the patent period. During the same period, the 1960s, availability of generic drug products began to increase significantly. The FDA at that time utilized the Abbreviated New Drug Approval (ANDA) process developed in the 1962 amendments for review and approval of generic drug products that were to be marketed after FDA approval and patent expiration of the originator new drug entity. Some legal questions, however, arose as to the use of the ANDA procedure for generic approval for both pre-1992 and post-1962 new drug approvals (see PATENTS AND TRADE SECRETS).

The Drug Price Competition and Patent Restoration Act of 1984 included two sections that addressed the use of the ANDA procedure and the loss of patent average time, since the review process mandated by the 1962 amendments (7). Title I of the 1984 Act governs the approval of NDAs, regardless of whether the originator's drug approval occurred before or after 1962. Title II covers the extension of patent coverage for a new drug entity. The Act permits a qualified owner (originator) of an approved new drug entity to apply for patent coverage extension. Such extension usually is approved for a time period equal to that defined as the regulatory review period, within certain limitations (7). This review period is generally accepted as one-half of the Investigational New Drug (IND) period, as authorized by the 1962 amendments, plus the post-IND preapproval period. If the applicant for such extensions has not exercised due diligence in seeking its NDA, the extension period can be reduced (7).

The world trade agreement, the General Agreement on Tariffs and Trade (GATT), resulted in a U.S. federal law, the Uruguay Round Agreement Act (URAA), that became effective in June 1995. Under this Act, numerous drugs are projected to gain months or even years of additional patent protection, depending on current patent expiration dates (8,9). The GATT provides new prescription drugs with 20 years of patent protection from the patent application date. The Drug Price Competition and Patent Term Restoration Act of 1984 can be a complicating factor. The U.S. patent office, as of June 8, 1995, has indicated that drugs that received patent extension by virtue of the 1984 laws would not receive further extensions under URAA GATT (8). FDA has indicated that pending ANDA applications for generic approval cannot be approved during the URAA extended periods (9). The generic drug industry opposes such a ruling, which would delay approval of drug products being reviewed through the ANDA process. Consumer groups have argued that such extensions significantly increase the cost of drugs to individuals and government financed programs. As of this writing (ca 1995), this issue needs clarification by U.S. courts (8).

The 1962 Amendments also mandated a review of safety and therapeutic efficacy for U.S. nonprescription, ie, over-the-counter or proprietary, products. There are an estimated 125,000–300,000 U.S. OTC products covering a variety of sizes, dosage form types, and dosage form strengths (4). The FDA has increased its approval rate for the switch of prescription drugs to nonprescription status in the 1990s. This procedure has gained impetus as more than 450 OTC products in 1994–1995 used ingredients and dosages only available by prescription in 1974–1975 (10).

The principal OTC pharmaceutical products include cold remedies, vitamins and mineral preparations, antacids, analgesics, topical antibiotics, antifungals and antiseptics, and laxatives. Others include suntan products, ophthalmic solutions, hemorrhoidal products, sleep aids, and dermatological products for treatment of acne, dandruff, insect parasites, burns, dry skin, warts, and foot care products (11). More recent prescription-to-OTC switches have included hydrocortisone, antihistamine and decongestant products, antifungal agents, and, as of 1995, several histamine H₂-receptor antagonists.

Personnel. A large number of personnel trained in a wide range of special skills are needed for the development of a new drug. Skills include organic synthesis, medical and analytical chemistry, microbiology and immunology, biochemistry, physiology, pharmacology, toxicology, and pathology (13). Likewise, in the development of safe, stable, and therapeutically effective drug products various physical chemistry principles apply and specialists trained in this phase of development, pharmaceuticals, assume such responsibility. These people become involved in the preformulation studies that investigate the properties of the new drug for inclusion in dosage forms, in the scale-up procedures that are needed to transfer dosage form preparation from laboratory batch sizes to manufacture batch sizes, and in the actual manufacture of the product. These specialists work closely with chemical engineers, especially during the scale-up phase.

Concepts and Processes. Contemporary dosage forms are drug delivery systems, designed and manufactured to achieve safe and effective therapeutic responses each time the forms are used as part of an appropriate regimen. Thus, the intent of the prescriber is accomplished when the product is used compliantly by the patient (12–14). Each drug product involves several interrelated concepts that must be considered in its design and manufacture (15). Examples include the following:

Component concept	Requirement
drug (active ingredient)	purity, stability, accuracy in measurement
nontherapeutic ingredients (excipients)	needed for safe and effective delivery of the active ingredient
unit process	procedures needed to ensure batch-to-batch, dose-to-dose reliability of safe and effective response
manufacturing technology	designed for patient compliance and product stability
packaging	to protect the drug product throughout its projected shelf-life
labeling	to ensure stability and safety
quality assurance procedures	efficacy
storage	

Attention to various physiochemical parameters of the drug moiety, such as particle size, crystalline form, and solubility, is vital to the design of a dosage form, as are its purity and accurate measurement. Nontherapeutic or excipient ingredients are selected to ensure stability (buffers, chelating agents (qv), antioxidants (qv), antimicrobial preservatives), and accuracy and precision of dosage (diluents, vehicles). Similarly, various types of excipients are used for specific types of dosage forms in order to permit their manufacture and desired therapeutic performances, eg, disintegrating agents for compress tablet formulations. Other excipients function as processing aids; for example, glidants that ensure effective flow of granulation during tablet compression. Lubricating agents are solids used in tablet compression to lubricate the die-walls and punch faces to prevent sticking, capping, and or excessive die-wall wear. Polymers find wide excipient use in dosage form design as viscosity-building agents in suspensions and emulsions and in the control of drug release in products prepared to achieve longer (8–12 h) than usual therapeutic periods. Various excipients are used to provide drug palatability for patients, eg, colorants (see COLORANTS FOR FOOD, DRUGS, COSMETICS, AND MEDICAL DEVICES) and flavoring agents (see FLAVORS AND SPICES) (15).

The selection of excipient ingredients is important. These must be both chemically and physically compatible with the drug moiety and cannot negatively affect product stability or therapeutic performance, ie, bioavailability. A comprehensive list of various types of excipient ingredients, with comment upon usage, is available (16).

The various preparation processes and technologies used in drug product manufacture also can effect product safety, stability, and performance, eg, compression during tablet manufacture. The principal processes used in dosage form manufacture are as follows (15).

Dosage form types	Processes
-------------------	-----------

liquid solutions	dissolution and filtration
parenterals	sterilization, lyophilization
liquid dispersion (suspensions, emulsions)	dispersion wetting of solids, homogenization
semisolid dispersions (ointments, creams)	levigation, melting
liquid semisolid capsules	solft gelatin encapsulation
suppositories	molding
solids (granules, capsules, tablets)	comminution, blending, granulation, compression, coating
aerosols	specialized packaging under pressure
general	heating, cooling, mixing

The therapeutically active drug can be extracted from plant or animal tissue, or be a product of fermentation (qv), as in the case of antibiotics. Frequently, it is synthesized and designed to correlate structure with therapeutic activity. Pharmacologic activity is first tested on laboratory animals. When the results are encouraging, physical and chemical properties are determined in the so-called preformulation stage, and analytical procedures are developed for quality control (see [QUALITY ASSURANCE](#), [QUALITY CONTROL](#)).

Biological characterization includes toxicological studies, dose relationships, routes of administration, identification of side effects, and absorption, distribution, metabolism, and excretion patterns. If the results are still acceptable, product formulation and dosage form are developed. The product should be pleasing to the patient and thus may contain flavoring and colorants.

Application for discovery and product patents must be made early in the process. Appropriate labels are designed and the product is submitted to the FDA for approval to begin human testing in the form of an Investigational New Drug Application (INDA). When such approval is granted, the clinical evaluation is developed as follows.

Phase I. This involves general testing for human pharmacology in healthy volunteers, ie, safe-dose adjustment; determination of absorption, metabolism, and excretion patterns; and monitoring for side effects. Usually fewer than 10 test subjects are involved.

Phase II. Initial clinical studies for therapeutic safety and efficacy are performed in volunteer patients who are suffering from the disease for which the drug has therapeutic promise. Recognition of toxic symptoms and side effects are vital at this point because these may occur here, even when not observed in animal studies or in Phase I.

Phase III. Drug samples are made available to select clinicians for use on large numbers of patients to obtain statistically significant data for safety and efficacy.

The INDA actually is a request for a Claimed Investigational Exemption to allow the transport of clinical samples of a nonapproved drug into interstate commerce for the testing in human subjects.

Manufacturing, analytical, and quality control procedures are thus established. Specifications for raw and in-process materials, as well as for final products per USP NF and in-house standards are also determined. Process and formula validation assures that each technological procedure in manufacture accomplishes its purpose most efficiently, eg, blending times for powdered mixtures in tableting, and that each formula ingredient is present in optimal concentrations (12). Thus, it serves to ensure process control (qv), reproducibility, and content uniformity.

Stability studies are developed to assure a desirable shelf-life period. These also establish limits of acceptability for impurities and degradation compounds, when present, and determine acceptable storage conditions for raw materials and the manufactured products. Stability studies are thus important to the determination of expiration dates for drug products.

In 1995, discussions among the United States, the European Community (EC), and Japan occurred to achieve harmonization of drug and drug product standards and to provide guidance to the worldwide pharmaceutical industry for acceptance of global regulatory filings. The International Committee on Harmonization (ICH) has proposed initial guidelines for the establishment of stability studies.

Finally, all data, including the results of the clinical investigation, are collected in a New Drug Application (NDA) and sent to the FDA. Once approved, the new drug goes into production. After manufacturing begins, the new drug products must be monitored in clinical use in the marketplace for reports of untoward reactions. This amounts to post-approval surveillance known as Phase IV. All such reports must be submitted to the FDA in a timely manner.

The FDA and USP have developed the Drug Product Problem Reporting Program to detect performance problems of marketed drug products. The pharmacist or physician can report any problems experienced with drug products and medical devices. In cases where the FDA and or manufacturer finds that a marketed product constitutes an actual or potential threat to the safety and welfare of the public, that product must be withdrawn from the marketplace, ie, recalled. Several classes of recalls exist, depending on the relative danger that the product exhibits. *Class I* drugs pose a serious health threat and may require withdrawal at the consumer level; *Class II* drugs pose a possible or potential health problem that usually means withdrawal at the pharmacy or wholesaler levels; and *Class III* drugs may present a remote hazard to health and safety.

U.S. Laws and Regulations Related to Drug Product Development and Manufacture. Until early in the twentieth century, drug products were made and sold in the United States having virtually no imposed control. Quality was generally poor. Many products were patent medicines of dubious value. Some were harmful and addicting.

The 1906 Food, Drug and Cosmetic Act (FD&C) addressed product contamination, but did not require premarketing approval of new drugs. In 1937, ethylene glycol, used as a vehicle for an elixir of sulfanilimide, caused more than 100 deaths. Thereupon, in 1938 the FDC act was revised to require advance proof of safety for new drugs, factory inspections, and various other controls. In 1941, the FD&C Act was amended to require the certification of safety of batches of penicillin and other antibiotics. The Durham-Humphrey amendments of 1951 established categories of drug products, ie, those dispensed by prescription only and those sold directly to the consumer. The former are often termed legend drugs, after the printed label legend that restricts availability; the latter are referred to as OTC products.

Other federal laws control the distribution of drugs that are subject to abuse, eg, the Harrison Narcotic Act of 1914, the Drug Abuse Control Amendment of 1965, and the Comprehensive Drug Abuse Prevention and Control Act of 1970. The last placed control for stimulants (qv), depressants, narcotics, and hallucinogenic drugs within the Bureau of Narcotics and Dangerous Drugs (BNDD) of the U.S. Department of Justice. A 1973 reorganization of that department created the Drug Enforcement Agency (DEA) (see [REGULATORY AGENCIES](#)).

The *United States Pharmacopeia* (USP) and the *National Formulary* (NF) are the recognized standards for potency and purity for most common drug products. The USP was first published in 1833; the NF in 1887. Upon adoption of the first Food and Drug Act in 1906 these compendia became official, ie, they publish the legal standards of quality, purity, and strength. The 1980 editions of USP(XX) and NF(XV) were combined for the first time; revisions take place every five years. USPXXIII NFXVIII is effective as of 1995.

After the thalidomide calamity in 1962, the FD&C Act was amended further (Kefauver-Harris amendments). A greater degree of safety testing was required before marketing, including attention to teratogenic potential. Specific phases of human clinical testing were established having strict controls (Phase I, II, and III), including patient consent and proof of therapeutic efficacy. In addition, evaluation after marketing is required, sometimes referred to as Phase IV.

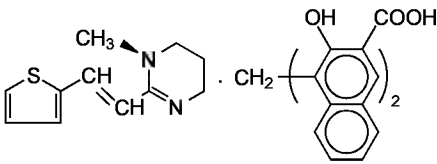
The concept of the NDA was introduced in the 1938 amendments for proof of safety of new drugs. The 1962 amendments added the concept of the Claimed Investigational Exemption (INDA) for a new drug. Thus, all the research and development data, including animal testing, for new drug substances or new dosage forms of an existing drug substance must be submitted to the FDA with the Investigational New Drug Application (INDA) for approval, before clinical testing. Approval of the INDA by the FDA grants the claimed Investigational Exemption which permits the introduction of the drug in question into interstate transportation for clinical human testing. The total research data (INDA plus clinical) comprise the NDA.

In 1966 the FDA utilized the services of the National Academy of Sciences–National Research Council (NAS–NRC) to establish the relative therapeutic efficacies of prescription drugs marketed between 1938 and 1962. Those products that were found to meet safety and efficacy requirements were allowed to stay on the market. Suitable changes were required for other products for compliance, ie, formulation or label changes, additional data

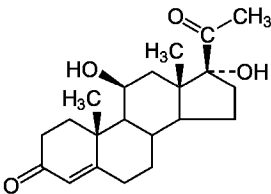
submissions, etc. Drugs that were in use prior to 1938 were approved without review, owing to the length of time of their safe and effective use.

The 1962 amendments also resulted in a massive review of nonprescription drugs toward the development of regulatory OTC monographs (17). The recommendations of review panels were first published in the *Federal Register* and the procedures are the same in the 1990s. Comments from the public are reviewed by the panel, which then issues tentative final monographs. After a second period of public comment and panel review, a final monograph must be developed which finally becomes regulation after another period of public scrutiny. Each stage in the process must be subject to review by the FDA. At any stage, concerned groups can petition for a reopening of the record in order to submit new information or make formal objections (18). Throughout the panel reviews, each drug is considered for recommendation in three categories: safe and effective for the indicated use (Category I), neither safe nor effective (Category II), and more data needed to prove the safety and efficacy (Category III). Until 1979, drugs that were placed into Category III status were allowed to be marketed for a time while the manufacturer provided additional data. However, at that point, a federal district court in Washington, D.C. ruled that once a final monograph is published, no Category III product may remain on the market (19).

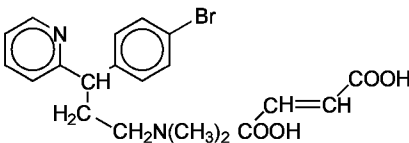
An outcome of the OTC review has been the switching of some prescription-only drugs to OTC status. Between 1972 and 1981, these included fluorides in mouth-rinse products, several vasoconstrictor drugs for anorectal therapy, the anthelmintic pyrantel pamoate [22204-24-6] (1), hydrocortisone [50-23-7] (2) for dermatologic products, the antihistamine brompheniramine maleate [980-71-2] (3), the bronchodilator methoxyphenamine hydrochloride [5588-10-3] (4), and two nasal decongestants, oxymetazoline hydrochloride [2315-02-8] (5) and xylometazoline hydrochloride [1218-35-5] (6) (19). As noted previously a significant number have also achieved this switch in status since 1981.



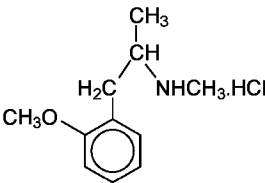
(1)



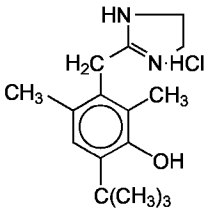
(2)



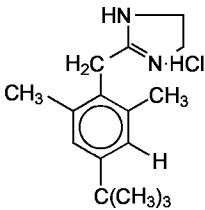
(3)



(4)

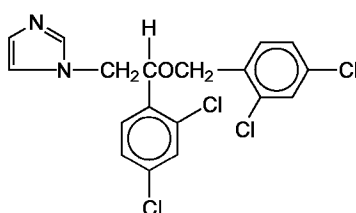


(5)

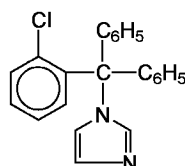


(6)

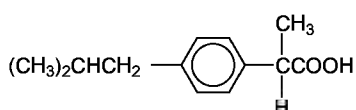
From 1981 to 1995 other drugs switched to OTC were the antifungal preparations containing miconazole nitrate [22832-87-7] (7) and clotrimazole [23593-75-1] (8); the antiinflammatory agent ibuprofen [15687-27-1] (9); and the histamine H₂-receptor antagonists famotidine [76824-35-6] (10) and cimetidine [51481-61-9] (11).



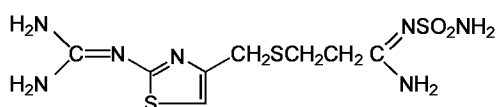
(7)



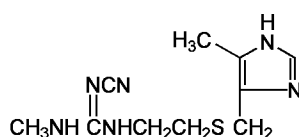
(8)



(9)



(10)



(11)

The Prescription Drug Marketing Act of 1988 amended the FD&C Act to establish newer safeguards for the national distribution of prescription drug products. It limits the reimportation of drug products manufactured in the United States to the product manufacturers only. It also prohibits the selling, trading, and purchasing of sample drug products and places restrictions upon their distribution.

Bioavailability, Bioequivalence, and Pharmacokinetics. Bioavailability can be defined as the amount and rate of absorption of a drug into the body from an administered drug product. It is affected by the excipient ingredients in the product, the manufacturing technologies employed, and physical and chemical properties of the drug itself, eg, particle size and polymorphic form. Two drug products of the same type, eg, compressed tablets, that contain the same amount of the same drug are pharmaceutical equivalents, but may have different degrees of bioavailability. These are chemical equivalents but are not necessarily bioequivalents. For two pharmaceutically equivalent drug products to be bioequivalent, they must achieve the same plasma concentration in the same amount of time, ie, have equivalent bioavailabilities.

For drugs approved originally between 1938 and 1962, the FDA has utilized the Abbreviated New Drug Application (ANDA) for review of generic products that are pharmaceutical equivalents of the initially approved products. In this way, costly duplication of animal and human experimentation is avoided. The new manufacturer has to show only that its manufacturing methodology, specifications, quality control, and labeling are acceptable. In some cases, the FDA does require proof of bioequivalence.

In 1980, the FDA first published "Approved Prescription Drug Products with Therapeutic Equivalence Evaluation" (20), which lists the products that have been approved under Sections 505–507 of the Federal Drug and Cosmetic Act. Supplements and annual updates are issued. "Current Good Manufacturing Practices" (CGMP) was first published by the FDA in 1963 and revised in 1973 as part of the Food and Drug Code of Federal Regulations (Title 21). It deals with all phases of pharmaceutical manufacturing and includes regulations covering clinical testing.

Bioavailability, important to the design and preparation of drug products, can be affected adversely by the selection of excipients and or the manufacturing processes used. Excessive pressure used in the compression of tablets, for example, could cause a tablet to pass through the gastrointestinal tract with no therapeutic effect. Figure 1, which represents the blood concentration–time curve for an orally administered tablet, illustrates the basic concepts involved in bioavailability determinations. The area-under-the-curve (AUC) represents the total amount of an administered dose that reaches the blood. The onset and duration of therapeutic activity can be influenced, as can the degree of untoward side effects (15).

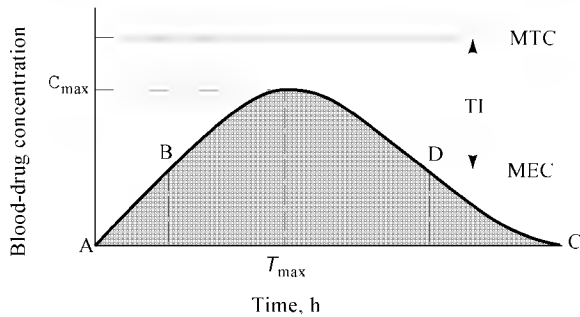


Fig. 1. Blood–drug concentration curve used to determine bioavailability and bioequivalence. C_{max} is the maximum drug concentration in the blood and corresponds to some T_{max} . The AUC (shaded) represents the total amount of orally administered drug; the time from points A to B represents drug onset, from points B to D, the duration; MEC = minimum effective concentration; MTC = minimum toxic concentration; and TI = therapeutic index.

Two similar dosage forms, eg, tablets, that contain the same amount of the same drug entity and meet USP NF and current good manufacturing practices (FDA) are referred to as pharmaceutical equivalents (PE). When, upon administration, such tablets achieve similar profiles of AUC, C_{max} , and T_{max} values (Fig. 1), these are adjudged to be bioequivalent. The FDA sets the degree of similarity needed to be so termed. This concept is of obvious importance for the acceptance of generic drug products for marketing.

Pharmacokinetics is the study of how the body affects an administered drug. It measures the kinetic relationships between the absorption, distribution, metabolism, and excretion of a drug. To be a safe and effective drug product, the drug must reach the desired site of therapeutic activity and exist there for the desired time period in the concentration needed to achieve the desired effect. Too little of the drug at such sites yields no positive effect ($<\text{MEC}$); too much ($>\text{MTC}$) leads to toxicity (see Fig. 1). For intravenous administration there is no absorption factor. Total body elimination includes both metabolic processing and excretion.

In cases of all but intravenous administration, dosage forms must make the active moiety available for absorption, ie, for drug release. This influences the bioavailability and the drug's pharmacokinetic profile. Ideally the drug is made available to the blood for distribution and elimination at a rate equal to those processes. Through technological developments drug product design can achieve release, absorption, and elimination rates resulting in durations of activity of 8–12 hours, ie, prolonged action controlled release drug products (21,22). Such products improve the compliance rate of drug usage by patients.

Manufacturing

Table 1 gives the common dosage forms of pharmaceuticals as of this writing (1995) including properties and uses (23). A comprehensive list and review of the varied types of dosage forms used in earlier years is also available (24). This treatise includes older forms of liquid preparations (decoctions, infusions, mucilages, fluid glycerates, brandies, essences, balsams, and oleoresins); solid preparations (medicated cones, cachets, dragees, insufflations, pills, and wafers); semisolid preparations (cataplasms, cerates, glycerogelatin, plasters, and vesicatories); and suppository-type preparations (bougies and pessaries).

Table 1. Pharmaceutical Dosage Forms

Dosage form	Constituents, properties	Uses
<i>Liquid solutions</i>		
aromatic waters	volatile solids or oils, water	flavoring agents, carminative action
liquors or solutions	water, chemicals	internally or externally
syrups	sweetener, solvent, medicinal agent	flavoring agent, medicinal
elixirs	sweetened hydroalcoholic soln, may be medicated	flavor or medicinal
spirits	alcohol, water, volatile substances	flavor or medicinal
tinctures	natural drugs, extracted with appropriate solvent; 10–20 g cm ^{3a}	external or internal
collodions	pyroxylin in ether, medicinal agent (castor oil, camphor)	external for corns and bunions
liniments	oily or alcoholic solutions, suspensions	external with rubbing
parenteral soln	sterile, pyrogen-free, isotonic, pH close to that of blood; oily ^b or aqueous suspension	intravenous, intramuscular, subcutaneous injection
ophthalmic	sterile, isotonic, pH close to that of tears; viscosity builder	eye treatment
nasal	aqueous, isotonic, pH close to that of nasal fluid; sprays or drops	nose treatment
otic	glycerol-based	ear treatment
mouthwash, gargles	aqueous, antiseptic	refreshment, short-term bacterial control
inhalations	administered with mechanical devices	medication of trachea or bronchioles
enemas, douches	aqueous soln or suspension, may include medicinal agent	irrigation of body cavity
<i>Liquid dispersions</i>		
suspensions	powder suspended in water, alcohol, glycol, or an oil; viscosity builders, wetting agents, preservatives	oral dosing, skin application
emulsions, lotions	oil-in-water (o w), or water-in-oil (w o)	oral, external or injection
gels, jellies, magmas	viscous, colloidal dispersions	internal (oral), external
gaseous solutions, dispersions	delivered in atomizers, nebulizers, aerosols, inhalers	external or internal
<i>Semisolid and plastic dispersions</i>		
ointments	hydrocarbon (oily), absorptive water-washable, or water-soluble bases; emulsifying agents; glycols; medicating agent	external
pastes and cerates	ointments with high dispersed solids or waxes, respectively	external
suppositories	theobroma oil, glycerinated gelatin, or polyethylene glycol base plus medicinal agent	insertion in body cavity
<i>Solids</i>		

powders		
bulk	comminuted or blended, dissolved in or mixed with water	external, internal
effervescent	CO ₂ -releasing base ingredients	oral
dusting	absorbents; lubricants	skin treatment
insufflations	insufflator propels medicated powder into body cavity	body cavities
lyophilized	reconstitution by pharmacist of unstable products	various uses, including parenteral and oral
capsules	small-dose bulk powder enclosed in gelatin shell; active ingredient plus diluent	internal
troches, lozenges	prepared by piping and cutting or disk candy technology; compounded with glycerogelatin	slow dissolution in mouth
tablet triturates	small molded tablets intended for quick complete dissolution, eg, nitroglycerin	oral (sublingual)
granules	particle size larger than powder	oral
compressed tablets	dissolved or mixed with water; great variety of shapes and formulations ^c	oral and external
pellets	for prolonged action	implantation
coated tablets	coating protective; slow release	oral

^a Concentrations of 1 g cm³ are called fluid extracts.
^b Repository dosage form.
^c Common excipients: diluents, disintegrators, binders, and lubricants (glidants).

Compressed Tablets. This popular type of dosage form offers convenience, stability, accuracy and precision, and good bioavailability of active ingredients. After the best formulation has been established, compressed tablets can be manufactured at high rates of speed on advanced equipment. Tablets can be made to achieve rapid drug release or to produce delayed, repeat, or prolonged therapeutic action (CONTROLLED RELEASE TECHNOLOGY, PHARMACEUTICAL).

Quality control during manufacture and of the final product assures batch-to-batch consistency and reliability. Bioavailability is checked in early batches produced for clinical testing. Other tests include uniformity of weight and contents, hardness (qv), disintegration rate, dissolution rate, and friability.

During the preformulation stage, the chemical and physical properties of the drug moiety are studied exhaustively to ensure stability, safety, bioavailability, and therapeutic efficacy. Tablets are produced directly by compression of powder blends or granulations, which include a small percentage of fine, particle-sized powders.

Granulation. Granulation methods can be wet or dry. Wet granulation cannot be used for drugs that are sensitive to moisture and heat. The powdered drug and diluent are blended with a dispersion of the binder excipient, eg, gelatin, to a consistency that can be screened to 840–1800-μm granules (10–20 mesh). These granules are dried on trays in hot-air ovens or fluid-bed dryers. The latter are more time efficient and can be modified by combining the granulating and drying stages. The dried granules are resieved, generally to 420–840 μm (20–40 mesh), and blended with powdered lubricants and disintegrants. In some instances a portion of the disintegrating agent can be added to the powder blend before the addition of the binder. The percentage of powdered ingredients, ie, the fines, should at this point be relatively small (ca 5%). The granulation blend is then compressed.

Dry granulation is used when the drug is not stable under the conditions of wet granulation and when the combined powders of a formulation cannot be compressed directly. One form, slugging, occurs when all the ingredients are blended and compressed on heavy-duty tablet presses. Generally, the pressure is greater than in regular compression and the resulting tablets or slugs are very large (2.5–5.0 cm dia) and weigh 20–30 g. These large, hard slugs are ground and screened to appropriate mesh size and then recompressed into final tablets. Chilsonation, another form of dry granulation, involves the use of roller compaction of the blended ingredients, followed by particle size reduction to appropriate sizes for compaction.

Direct Compression. This process is relatively simple and time saving. All the ingredients are blended and then compressed into the final tablet. This is an excellent method, but encumbered by a number of problems. Not all substances can be compressed directly, necessitating a granulation step. Likewise, the flow properties of many blends of fine, particle-sized powders are not such as to ensure even filling of the die cavities of tablet presses. In addition, air entrapment can occur.

The availability of spray-dried lactose, microcrystalline cellulose, and other excipients allows for the use of granular rather than powdered phases. This eliminates some of the problems of particle segregation according to size (demixing) and even flow to the die. Direct compression eventually may be the preferred method of tablet preparation.

Tablet Press. The main components of a tablet compression machine (press) are the dies, which hold a measured volume of material to be compressed (granulation), the upper punches which exert pressure on the down stroke, and the lower punches which move upward after compaction to eject the tablets from the dies. Mechanical components deliver the necessary pressure. The granulation is fed from a hopper with a feed-frame on rotary-type presses and a feeding shoe on single-punch presses. A smooth and even flow ensures good weight and compression uniformity. Using the proper formulation, demixing in the hopper is minimized.

The actual compression process is a cycle of die fill, compaction by intervention of the upper punch using great pressure on the granulation material in the die, and upward movement of both punches to achieve ejection of the tablet from the die. Single-punch presses have only one die-and-punch arrangement and the compression is quick, with little dwell time of the top punch in die.

Rotary tablet presses can accommodate many punch–die units. The dies are set in a rotating, circular, metal table and the punches ride in appropriately designed cam tracks or channels in the head and foot areas of the press to achieve the necessary upward and downward stroking action. The central shaft mechanism drives these rotating components in synchrony, producing the designed number of tablets in each cycle. The granulation is fed from the hopper to the dies, passing below the feeder frame at a point when the lower punches are in their lowest position. The frame may contain some devices, such as rotating spindles, to induce or force granulation into the dies, as a means to ensure more accurate and uniform fills. Pressure-release devices allow a lift release if an overload at the die occurs.

Single-punch machines produce approximately 100–150 tablets per minute. Depending on numbers of die per punch units, standard rotary presses can produce 5000 tablets min, and even more with a double-sided rotary press. The newest high speed presses can achieve 12,000 tablets min.

Some presses are equipped with strain gauges at key points in the overall feed–compress–eject cycle. Thus, these measure compression and ejection forces. Tight specifications for punch lengths and well-designed and prepared granulations have led to better control of variations in tablet weight. In fully automated presses, weight variations are adjusted by computer.

Compressed tablets that are composed of several layers require specially adapted presses designed with several fed hoppers. For a two-layer tablet, one granulation is first fed to a die and partially compressed into a soft tablet. The second granulation is added, and the total die components then are compressed fully. Such procedures are used when the tablet ingredients may be incompatible, which requires separate granulations. If needed, a layer of inert ingredient, eg, lactose, is inserted between the two.

Layered tablets are also used for a prolonged or sustained therapeutic effect. In this case, one layer disintegrates and dissolves rapidly to provide the initial dosing, whereas the other is designed for controlled release.

Formulation. Compressed tablet formulations contain several types of inert, adjuvant ingredients necessary for proper preparation and therapeutic performance. Tablets designed to be swallowed need diluent, disintegrating, binding (adhesive), and lubricating inert ingredients, whereas

troches or lozenges intended to be dissolved slowly in the mouth should not disintegrate quickly, need more binder, and no disintegrant. Lactose or dicalcium phosphate are common diluents, whereas starch and cellulose derivatives are used as disintegrating agents. Sublingual tablets are designed to dissolve rapidly under the tongue to provide rapid absorption as in the case of nitroglycerin usage.

Glidants are needed to facilitate the flow of granulation from the hopper. Lubricants ensure the release of the compressed mass from the punch surfaces and the release ejection of the tablet from the die. Combinations of silicas, corn starch, talc (qv), magnesium stearate, and high molecular weight poly(ethylene glycols) are used. Most lubricants are hydrophobic and may slow down disintegration and drug dissolution.

Colors and flavors increase the elegance and acceptability of the product. Sometimes colors are used for identification.

Effervescent tablets disintegrate by virtue of the chemical reaction occurring in water between component ingredients, such as sodium bicarbonate and citric or tartaric acid, to achieve release of carbon dioxide. Interest exists in the pharmaceutical industry for technologies to produce fast dissolving tablets and various concepts are being investigated. One is the use of mildly effervescent bases and adaptation of technology similar to that used to make cotton candy. Another, intended to produce a product to dissolve very quickly in the mouth, involves preparation of a lyophilized porous wafer containing a therapeutic substance that dissolves in seconds in saliva, when placed on the tongue. This technology (Zydis System) is being marketed to pharmaceutical manufacturers by its developer, R. P. Scherer Co.

Coating. Sugar or film coatings offer protection from moisture, oxygen, or light and mask unpleasant taste or appearance. Enteric coatings delay the release of active ingredients in the stomach and may prolong the onset of therapeutic activity. The latter are used for drugs that are unstable to gastric pH or enzymes, cause nausea and vomiting or irritation to the stomach, or should be present in high concentration in the intestines, eg, preoperative sterilization of the gut or as anthelmintics. Effectiveness depends on the varying pH patterns of the gastrointestinal tract and the enzymes present for dissolution and aqueous solubility.

Enteric coating is also used for repeat-action tablets, which contain an enteric-coated core tablet and a sugar or film-coated second dose, permitting the administration of two doses simultaneously. The core dose is released several hours after the initial, outer dose.

Some tablets that provide a sustained period (up to 8–12 h) of therapy may be coated during processing. A portion is released first to bring the drug to the desired blood concentration (onset of activity), whereas a sustained-release portion maintains an effective level for a prolonged period of time (duration of activity), eg, by coating erosion or diffusion of drug through it.

Sugar Coating. Sugar coating is applied in rotating, pear-shaped or short cylindrical pans. The cores are usually somewhat harder than in uncoated tablets to withstand the rigors of tumbling. They are first dusted and then wetted with a solution of concentrated sucrose, gelatin, acacia, or methylcellulose that imparts adhesiveness to the surfaces. Next, powdered sugar is applied with continued tumbling. The batch is then dried with warm air. This alternative subcoating procedure that uses syrup and powder is continued until the tablet is rounded. The core tablet may need to be sealed with a thin coat of pure shellac before it is subcoated. Color may be added to the smoothing coats. Finally, the tablet is polished with a waxy composition such as carnauba wax.

Sugar coating is time-consuming, requires skilled operators, and increases the tablet weight, sometimes to twice that of the uncoated core.

Film Coating. Film coating in pans is a much quicker procedure than sugar coating. The coating is much thinner and the process is easily automated or programmed. Various polymer solutions (eg, cellulose derivatives) are used that form films upon drying. Plasticizers improve film flexibility. The polymer may be water soluble or produce an enteric effect. Until the late 1980s, organic solvents such as acetone were used. However, because of increased cost and disposal problems of such solvents, aqueous-based solutions or dispersions of the polymers have become popular.

Gelatin Coating. A more recent development in tablet coating involves the use of gelation as the coating material to produce geltabs. If a tablet is compressed as a capsule-shaped unit prior to gelatin coating it is called a gelcap. Such tablets are dipped into a reservoir of a molten gelatin mixture, similar to the production of empty, hard gelatin capsule shells. The gelatin coating facilitates swallowing.

Air-Suspension Coating. The Wurster process utilizes a cylindrical chamber in which the cores are suspended in a controlled stream of air. Film coatings are applied by introducing the coating solution into the airstream, where the solvent evaporates quickly. The process is much quicker than film coating; however, care must be taken to avoid destruction of the cores by attrition in the air stream.

Compression Coating. In this dry process an outer coating is compressed around a core tablet, producing a tablet within a tablet. This requires sophisticated tableting presses that cannot be run at high speeds. The granulations are similar to those for uncoated, compressed tablets. This procedure is employed for drugs that are sensitive to moisture, oxygen, or light. Likewise, if two active incompatible ingredients are present, one can be granulated in the core and the other incorporated in the coating. Pressure coating also can be used to achieve prolonged periods of therapeutic effect. For example, the core can be formulated as a slowly eroding matrix that provides the sustained portion of therapy, whereas the initial dose is included in the compression coat.

Capsules. Capsules are made in two types. In hard-gelatin capsules, powders or granules are enclosed in rigid gelatin shells. Soft-gelatin capsules contain glycerol as well as gelatin and maintain plastically even when dried. Hard-gelatin capsules are made in two sections, cap and body, which are then filled, whereas soft-gelatin capsules are formed and filled in succession in one manufacturing procedure. Soft-gelatin capsules are generally filled using nonaqueous solutions, although powders can also be used. Most drug companies buy the hard-gelatin shells from external sources. These are made by dipping precisely tooled pins into controlled solutions of gelatin. A film of gelatin adheres to the pins. Upon drying, the units are trimmed to specified length, removed from the pins, and the cap and body portions are joined. Various colors can be incorporated (see GELATIN).

The formulations of filled, hard-gelatin capsules are generally less complex than those of compressed tablets, and require no binders or disintegrators. Upon swallowing, the capsule shell dissolves quickly and the powder ingredients are available for dissolution. Because no initial disintegration step is needed, bioavailability of drugs in capsule formulations is generally better than that of compressed tablets. The capsules are filled by various high speed machines. Occasionally the pharmacist has to perform this procedure manually.

The size system of capsules is inversely related to the volume. A No. 1 capsule is larger in volume than a No. 2; a No. 0 is larger than a No. 1. For human consumption, No. 0–2 are most common. Hard-shell gelatin capsules vary in size from those that contain 100 mg of drug to those for veterinary use, which contain several grams.

For prolonged action therapy, granular-sized encapsulated particles, ie, beads, are used and can be both uncoated or coated. The uncoated beads provide the initial dose; the others are made to dissolve at various rates depending on the coating type and thickness.

Hard-gelatin capsule shells, prepared from gelatin and water in various sizes, are made of two component parts, base and body. They can be compounded extemporaneously by pharmacists or manufactured by high speed machinery. Soft-gelatin capsules contain glycerin, as well as water and gelatin. These cannot be extemporaneously prepared owing to the gelatin character. Rather, these require heating process and sealing. As shown in Figure 2, for the soft-gelatin capsules two rolled sheets of gelatin are kept appropriately softened using minimal (40°C) heat. The rolled sheets meet in an injection wedge where the drug mix (solution; dispersion) is injected into its ultimate form upon sealing. Various shapes and sizes are possible by mold adaptations.

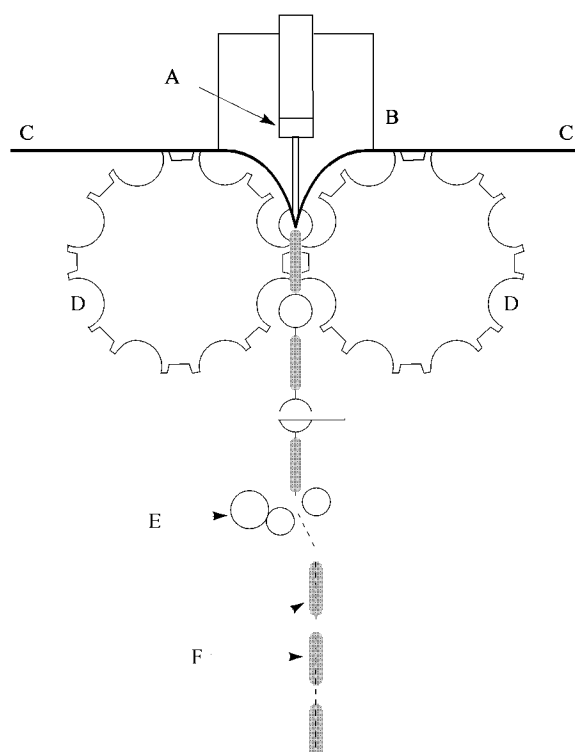


Fig. 2. Rotary die process of soft gelatin capsules where A represents the drug mix; B, the mold wedge; C, gelatin ribbons; D, die rolls; E, capsules; and F, scrap gelatin. Courtesy of R. P. Scherer Corp.

Prolonged Action/Controlled Release, Orally Administered Solid Dosage Forms. The therapeutic purpose of prolonged action and controlled release solid, oral drug products is to maintain safe and effective concentrations of the drug in the blood for 2–4 times longer than those times achieved using regular compressed tablets or capsules. This is accomplished by releasing one portion of the drug quickly, whereas the remaining portion is released at a rate that approaches the elimination rate. Ideally, the second portion should be released at a zero-order rate to achieve this profile. The technologies used for such controlled release only approach such a rate, but do accomplish the increased therapeutic period. These oral products mainly use diffusion-controlled or dissolution-controlled release profiles. The more recognized technologies used to achieve these methods include ion-exchange (qv) resins, coated micropellets, barrier coatings (see **BARRIER POLYMERS**), drug embedment in either slowly eroding or plastic matrices, swelling hydrogels of various polymer resins, drug complexation, and osmotic pressure controlled tablets (21,22). Other technologies that have been attempted or tested include altered density micropellets, prodrugs, and bioadhesives.

The best drug candidates for incorporation into prolonged action systems are uniformly absorbed throughout the gastrointestinal (GI) tract, have medium (2–8 h) biological half-lives, and are prescribed for chronic maintenance use. Drugs in large doses are difficult to formulate into such products (see **CONTROLLED RELEASE TECHNOLOGY, PHARMACEUTICAL**).

Since the development of the Spansule brand (SmithKline Beecham) of coated beads and granules in the late 1960s, various drug product technologies have been developed and patented to achieve extended durations of therapeutic effects. Each of these does so by various mechanisms of control of drug release from administered dosage forms. Each method has its advantages and disadvantages, a discussion of which is available in the pharmaceutical literature (see **DRUG DELIVERY SYSTEMS**) (21).

Coated Beads or Granules. Coats of varying thickness are applied to beads containing the appropriate amounts of drug (see **MICROENCAPSULATION**). The rate of dissolution depends on the rate of dissolution or disintegration of the coating and thus varies with the thickness of the coating. Various proportions of such coated beads are incorporated in gelatin capsules or compressed tablets together with appropriate amounts of uncoated beads to initiate onset of activity.

Eroding of Slow-Releasing Core Tablets. The sustained-dose portion of a drug is granulated with hydrophobic materials such as waxes, fatty acids, or fatty alcohols and compressed into a core. The initial dose is added to the core by a modified sugar coating process or by compression coating. Thus, a tablet within a tablet is created. The core erodes slowly to release the active ingredient.

Leaching from Carriers. The drug is granulated with inert plastic resins and water-soluble channeling agents. This mixture is compressed to form a porous plastic tablet with drug and channeling agent(s) entrapped in many veins or channels. The water-soluble excipient, ie, the channeling agent, attracts water from the gastrointestinal tract to the drug in the channels. The drug dissolves and passes into the gastrointestinal fluid. Such a leaching effect apparently occurs at rates suitable to accomplish a rapid initial onset of activity followed by a prolonged period of therapy. The exhausted plastic core is excreted.

Ion-Exchange Resins. Cationic-exchange resins which are combined with an appropriate form of the drug for which sustained activity is desired may be used. In the gastrointestinal tract the drugs ions are exchanged for other ions. Using the appropriate ion-exchange resin, the drug is delivered to the gastrointestinal fluids at a rate that produces sustained therapeutic activity. The rate of release also depends on the size of the bead resins. Thus, smaller beads have larger total surface area, the drug is exchanged at a faster rate, and the activity is less prolonged. Anionic-exchange resins have also been used.

Tannate Complexation. Certain drugs, those that contain amine groups, complex readily with tannic acid. Such complexes release the drug gradually and uniformly. The rate seems to be affected by the pH and the electrolytes present in the gastrointestinal tract. At lower pH, the drug is released more quickly. Other complexing compounds have also been used.

Hydrogels. Controlled swelling of hydrophilic polymers, derived from the glossy rubbery properties of polymers, is used to control the rate of drug release from matrices. In the rubbery state, accomplished by lowering the polymer's glass-transition temperature to an appropriate level, the dispersed drug diffuses as the polymer swells in the presence of water.

Osmotic Pressure Controlled Oral Tablets. Alza Corp. has developed a system that is dependent on osmotic pressure developed within a tablet. The core of the tablet is the water-soluble drug encapsulated in a hydrophobic, semipermeable membrane. Water enters the tablet through the membrane and dissolves the drug creating a greater osmotic pressure within the tablet. The drug solution exits at a zero-order rate through a laser drilled hole in the membrane. Should the drug itself be unable to provide sufficient osmotic pressure to create the necessary pressure gradient, other water-soluble salts or a layer of polymer can be added to the drug layer. The polymer swells and pushes the drug solution through the orifice in what is known as a push-pull system (Fig. 3). The exhausted drug unit then passes out of the body in fecal matter.

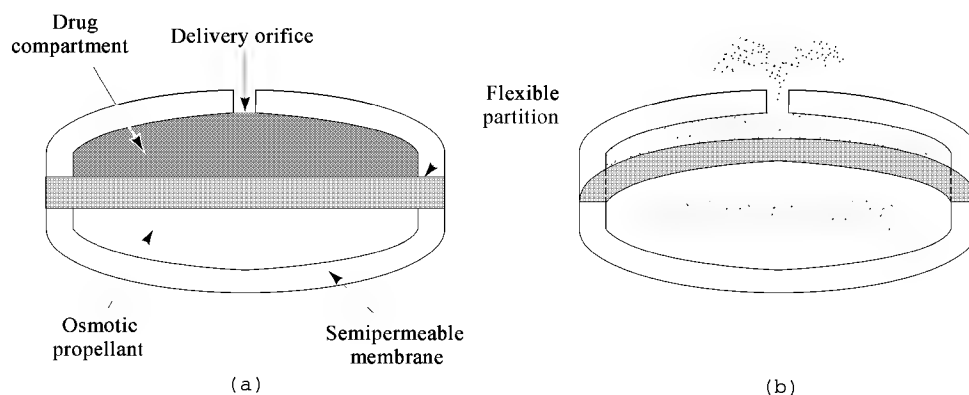
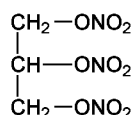


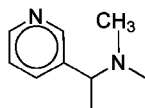
Fig. 3. (a) Cross section of the push-pull oral osmotic system (OROS), which has an inner flexible partition to segregate the osmotic propellant from the drug compartment. (b) Push-pull OROS in operation with the propellant imbibing water, increasing in volume, and pushing the drug out of the device through the delivery orifice (25).

This technology is relatively expensive to produce. Special excipients and equipment, such as a laser unit to drill the necessary hole for drug release, are required. However, the achievement of very steady blood levels of a drug for sustained periods, ie, zero-order rate release, of therapy is advantageous.

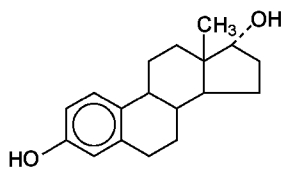
Other Prolonged Action Drug Products. The USP NF recognizes several nonorally administered, prolonged action controlled release drug delivery systems including transdermal, ocular, and intrauterine systems (23). The transdermal systems include medicated adhesive patches of various types. One patch technology utilizes a drug reservoir from which the drug diffuses through a rate-controlling membrane to and through the skin (see MEMBRANE TECHNOLOGY). Another type involves embedment of appropriately coated drug pellets into the adhesive of the patch. Nitroglycerin [55-63-0] (12), nicotine [54-11-5] (13), estradiol [50-28-2] (14), scopolamine [51-34-3] (15), and fentanyl [437-38-7] (16) are drugs that have been developed into such dosage forms.



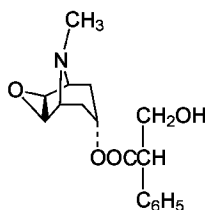
(12)



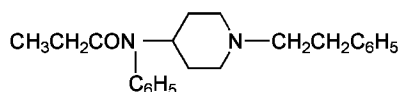
(13)



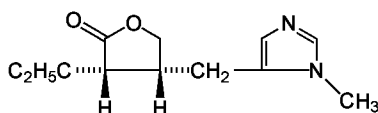
(14)



(15)



(16)



(17)

Ophthalmic drug delivery systems (qv) have been developed to deliver controlled drug quantities for a prolonged time (up to seven days) to the eye, eg, pilocarpine [92-13-7] (17). Alza Corp. in conjunction with Ciba-Geigy Corp. originally marketed such a product known as Ocusert to treat glaucoma.

This system utilizes specific membranes, between which the drug reservoir is enclosed (Fig. 4). A tiny elliptical disk, inserted into the cul-de-sac of the eye, releases pilocarpine steadily. The drug is delivered through selected polymeric membranes. The drug reservoir maintains a saturated solution between the membranes which acts osmotically as the driving force for the drug to diffuse through the rate-limiting membranes.

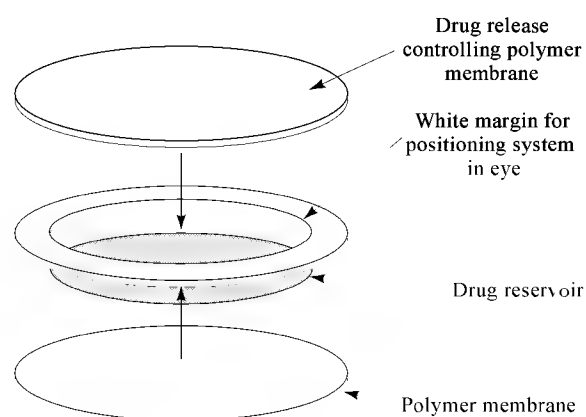


Fig. 4. Components of the Ocusert Pilo-40 therapeutic system (25).

Alza Corp. has also developed an intrauterine device, Progestasert, designed to release progesterone [57-83-0] by diffusion through a rate-controlling membrane for up to one year. The drug reservoir is built into a T-shaped device that is inserted intravaginally (15).

A prolonged action controlled release system developed to deliver levonorgestrel for contraceptive therapy involves implantation of a set of flexible closed capsules made of demethylsiloxane-methylvinyl-siloxane copolymer (see CONTRACEPTIVES). Each capsule measures 2.4 mm in diameter and 34 mm in length. A set of six such capsules is surgically implanted beneath the skin of the upper arm. These capsules are intended to be removed by the end of the fifth year after implantation.

Repeat and Delayed Action Oral Dosage Forms. Repeat action tablets provide prolonged periods of therapy, usually twice that of conventional release tablets, eg, eight hours instead of four. These are designed to release two portions of drug. The first portion is incorporated in an outer shell and is released entirely at one time, as for conventional tablets. The second portion is incorporated in a coated compressed core unit. The coating is designed to erode or dissolve at a rate such that the core dose is completely released when the concentration of the drug in the blood from the first portion approaches the MEC, after reaching its peak concentration. However, the drug release is not controlled to be gradual over the eight hours.

Delayed action solid products are designed like conventional dosage forms to release all their drug contents at one time, but only after a delayed period. Thus, the duration of action and the blood concentration-time curve is like that of a conventional product. However, the onset time is purposely designed to be long.

Such products are generally used when the drug is nauseating, irritating to the stomach, or chemically degraded by stomach pH and/or enzymes. In such cases, coatings that do not erode or dissolve in the stomach, but do so in the small intestine (enteric coatings), are used. An example of such a coating ingredient is cellulose acetate phthalate.

Liquid Dosage Forms. Simple aqueous solutions, syrups, elixirs, and tinctures are prepared by dissolution of solutes in the appropriate solvent systems. Adjunct formulation ingredients include certified dyes, flavors, sweeteners, and antimicrobial preservatives. These solutions are filtered under pressure, often using selected filtering aid materials. The products are stored in large tanks, ready for filling into containers. Quality control analysis is then performed.

Dosage forms of naturally occurring materials having therapeutic activity are prepared by extractive processes, especially percolation and maceration. Examples of such dosage forms have included certain tinctures, syrups, fluid extracts, and powdered extracts.

Solutions for external or oral use do not require sterilization but generally contain antimicrobial preservatives. Ophthalmic solutions and parenteral solutions require sterilization (qv).

For the preparation of suspensions and emulsions, colloid mills and homogenizers, respectively, are used. Ultrasonic mills that utilize vibrating reeds in restricted chambers to reduce the particle size of the dispersed ingredients can also be employed (see COLLOIDS; ULTRASONICS).

Semisolid Dosage Forms. The ingredients that constitute the base of ointments, eg, petrolatum and waxes, are melted together, powdered drug components are added, and the mass stirred with cooling. Generally, the product then is passed through a roller mill to achieve the particle-size range desired for the dispersed solid. Pastes are ointments having relatively large, dispersed solid content, and are prepared similarly.

Creams are semisolid emulsions either water-in-oil (w/o) or oil-in-water (o/w). Generally, the ingredients of the two phases are heated separately to ca 70–80°C. The phases are then mixed and stirred vigorously to achieve emulsification. Such stirring is continued until the product has been cooled sufficiently. For further reduction of the internal-phase droplets, the product may be passed through a homogenizer before final cooling. A solid ingredient can be added to the appropriate phase before emulsification or may be dispersed at some point after the emulsification step.

Suppositories. These semi-rigid, plastic dosage forms are designed to deliver a unit dose of medication to body cavities, ie, rectum, vagina, or urethra. Depending on the base, suppositories either melt (cocoa butter) at body temperature or dissolve (poly(ethylene glycol)s, glycerogelatin) in the fluids of the cavity. They can be used for systemic therapy (rectal suppositories) or for localized treatment. Rectal suppositories are a route of administration in comatose conditions or after gastrointestinal surgery, and for pediatric patients.

Cocoa butter-based suppositories can be prepared manually by pharmacists by mixing the ingredients to a pliable consistency in a mortar. This mass is then rolled into a thin cylinder and cut into units that represent one dose each. Melting the ingredients together and molding them into appropriate units in metal or rigid plastic molds generally is preferred. Formulations utilizing poly(ethylene glycol) or glycerogelatin bases must be prepared by molding because of the character of the individual base ingredients. On a large scale, suppositories are produced by molding.

Parenteral Dosage Forms. The most commonly used forms for drug products designed and manufactured for injection through the skin include those meant for subcutaneous, intramuscular, and intravenous administration (15,26–29). Other types include intradermal, intraarticular, intrathecal, intraspinal, intracisternal, and intraocular. Such dosage forms generally are termed injections and can be grouped into several categories: solutions ready for injection; powdered, soluble ingredients in appropriate containers that are combined with an appropriate solvent prior to use; suspensions that are ready to be combined with a vehicle prior to use; and emulsions (qv) (26,27). The route of administration and the physical nature of the injection have direct bearing on the selection of parenteral therapy. Injectable suspension (USP NF) for example, should not be administered intravenously because of the inherent danger of the suspended solid particles (26).

Intravenous aqueous injections provide an excellent means of achieving a rapid therapeutic response. Parenteral product design, eg, vehicle and other excipient selection, as well as choice of route of administration, can prolong therapeutic activity and increase onset times. Thus, oily solutions, suspensions, or emulsions can be administered by subcutaneous or intramuscular routes to create prolonged effect, ie, depot injection (28).

Several factors of design and manufacture are of great importance: sterility, absence of pyrogens and foreign particulate matter, and tonicity. The last, when adjusted to the osmotic pressure of body fluids in the case of aqueous solutions, reduces the risk of tissue irritation and pain. The USP NF utilizes the designation "Large Volume Intravenous Solution" for single-dose injections for iv use packaged in containers containing more than 100 mL. "Small Volumes" injections are those iv solutions containing 100 mL or less (27).

The USP recognizes three forms of water for parenteral dosage forms. Water for injection is prepared by reverse osmosis or distillation, which

removes nonvolatile pyrogens. It contains no added substances and is intended for solvent use in the preparation of parenteral solutions. When the solution is prepared under aseptic conditions, ie, not sterilized by filtration or in the final container, such water must be sterilized and protected from microbial contamination.

Sterile water for injection is used mostly for the solution or suspension of drugs just before injection. In containers of 30-mL capacity or less, it may contain a bacteriostatic agent. Inclusion of such agents in larger volumes can cause toxicity.

Bacteriostatic water for injection is sterile and pyrogen-free and contains bacteriostatic agents. The drug involved must be compatible with the antimicrobial agents present.

In addition to these forms of water, several other official aqueous vehicles can be used. These are isotonic injections that can be sterilized, eg, sodium chloride, Ringer's, dextrose, dextrose and sodium chloride, and lactated Ringer's. Addition of water-miscible solvents such as ethanol or propylene glycol increases solubility and stability.

Some fixed oils, such as cottonseed oil or peanut oil, and esters, eg, isopropyl myristate, may be used as solvent systems for parenteral drugs. Mineral oil and paraffins should not be used, because these are not metabolized and may irritate tissue. Various other additives are needed for stability, sterility, and isotonicity: antimicrobial preservatives, antioxidants (qv), chelating agents (qv), and buffers. No parenteral container material is completely inert to parenteral solvent systems.

Plastic components can be leached into the product and the alkalinity also can be affected by certain types of glass (qv). Particulate matter can be introduced by flaking from container surfaces. The containers also must be able to withstand the heat and pressure of sterilization.

Containers should be clear in order to allow detection of foreign particles. Outer coverings minimize irradiation. Plugs used as stoppers are selected with care to prevent flaking into the contents and possible component leaching into the product.

Traditionally, glass has been the preferred container material. The USP has adopted a classification of glass types acceptable for drug container use: Type I, borosilicates glass; Type II, a soda–lime treated glass; Type III, a soda–lime glass; and NP (nonparenteral), a soda–lime glass that is not suitable for parenteral products. There are two official USP tests: the powdered glass and the water attack test. In general, Type I glass is preferred; it is expensive, however. Types II and III may also be used for parenteral products.

Increasingly, plastics are being used as parenteral packaging (qv) materials. Plastics such as poly(vinyl chloride), polyethylene, and polypropylene are employed. However, plastics may contain various additives that could leach into the product, such as plasticizers (qv) and antioxidants. Permeability of plastics to oxygen, carbon dioxide, and water vapor must be tested in the selection of plastic containers. Furthermore, the plastic should withstand sterilization. Flaking of plastic particles should not occur and clarity necessary for inspection should be present.

Rubber is a popular closure component, and additives such as vulcanizers, pigments (qv), or antioxidants may leach into the product. In cases where rubber closures are penetrated by needles in dosing, bits of the closure (coring) could enter the product. Thus, such closure components must be sufficiently tested before use.

Commercially available containers for use with parenteral products include single-dose ampuls that are heat sealed and opened by snapping at the point of least diameter, vials for multidose use, and bottles and pliable bags that are used for large volumes such as needed in intravenous infusions. Container size can vary from 1 mL to 1 L. Generally volumes up to 100 mL are available as ampuls or vials.

Parenteral products are sterilized in containers soon after packaging by dry or moist heat under pressure (autoclaving). Drug solutions that are degraded chemically by heat can be sterilized by filtration through bacteria-retaining filters into sterilized containers under aseptic conditions, and then aseptically closed. Needles, syringes, and administration sets are sterilized with gas, ie, ethylene oxide. Ionized radiation has been used to sterilize sutures, dressings, needles, etc. However, gaseous and radiation sterilization are not suitable for liquid preparations.

Aseptic techniques must be scrupulously followed throughout the packaging stages of parenteral production. Sources of contamination are controlled strictly by using laminar-flow hoods having high efficiency particulate air (HEPA) filters. The availability of this technology, a spin-off of early aerospace research, has led to injectable products being prepared in hospital pharmacies. Ophthalmic solutions also can be prepared by this technique using bacterial filtration.

The industrial areas used for parenteral production call for careful design and conscientious maintenance of an aseptic environment (see CONTAMINATION CONTROL (SUPPLEMENT)).

Prolonged Action Parenterals Injections. Intramuscular injections have been developed to achieve prolonged therapeutic effects. This can be accomplished by suspension of drug particles in oils or flowable gels, from which the drug slowly diffuses. Aqueous suspensions can also provide such therapeutic response. In these cases, the solid drug crystals generally are quite water insoluble and of a controlled particle size and crystallized form. An example of such a product is Sterile Medroxyprogesterone Acetate Suspension used for its contraceptive property. Such an injection is designed to provide up to three months of contraceptive activity. Another such product is a depot injection of leuprolide acetate, an analogue of gonadatropin-releasing hormone (see DRUG DELIVERY SYSTEMS). In this case, the product is a sterilized powder of microspheres to be suspended upon the addition of an appropriate diluent and intended for monthly injection.

Lyophilization. Lyophilization is essentially a drying technology. Some drugs and biologicals are thermolabile and or unstable in aqueous solution. Utilization of freeze drying permits the production of granules or powders that can be reconstituted by the addition of water, buffered solution, or mixed hydrophilic solvents just prior to use, eg, certain antibiotic suspensions.

Initially, the product to be made using lyophilization is prepared as an aqueous solution or suspension, which is then cooled rapidly to a predetermined temperature. Such temperature is below the eutectic point and generally approaches -50°C. The freezing chamber is sealed and the frozen material subjected to heat under high vacuum conditions. The liquid portion sublimates, leaving the desired solid drug or biological. The process continues until less than 1% moisture remains in the dried components. Reabsorption of moisture can occur, necessitating quick removal from the freezer chamber into appropriate containers in a low humidity environment. When the lyophilized product is to be prepared for parenteral use, sterile conditions are maintained throughout the process. The dried drug or biological residue is porous upon sublimation of the ice crystals. Such surface character increases its rate of dissolution.

Ophthalmic Dosage Forms. Ophthalmic preparations can be solutions, eg, eye drops, eyewashes, ointments, or aqueous suspensions (30). They must be sterile and any suspended drug particles must be of a very fine particle size. Solutions must be particle free and isotonic with tears. Thus, the osmotic pressure must equal that of normal saline (0.9% sodium chloride) solution. Hypotonic solutions are adjusted to be isotonic by addition of calculated amounts of tonicity adjusters, eg, sodium chloride, boric acid, or sodium nitrate.

Single-dose preparations intended for use in eye surgery do not contain excipient ingredients, in order to avoid tissue irritation. However, multiple-dose containers may require antioxidants (qv), antimicrobial preservatives, or buffers to maintain stability and sterility. Such solutions are packaged in polyethylene flexible dropper units called droptainers or in glass dropper bottles.

Ophthalmic ointments usually contain petrolatum as the base. The petrolatum is sterilized by dry heat and combined with the sterile drug powder under aseptic conditions. Ophthalmic suspensions contain very fine (~ 10 μ) particle sized solids suspended in an aqueous vehicle. The vehicle is adjusted to isotonicity and viscosity-increasing excipients, chelating agents, and surfactants also may be needed. The aqueous vehicle in these cases is generally autoclaved and mixed with sterile drug powder aseptically (30).

Radiopharmaceuticals. Radioactive isotopes for human use in the diagnosis and treatment of disease states are called radiopharmaceuticals (qv) (15,31). Whereas the dosage form types used, eg, solutions or injections, are traditional, special handling of these products during compounding, transport, and use is vital. Most are administered intravenously and shortly after preparation. Examples of drugs used in such products include ⁵⁷Co-cyanocobalamin, ¹²³I-sodium iodide, ²⁰¹Tl-thallium chloride, ^{99m}Tc-technetium, ¹³¹I-sodium iodohipparate, and ³²P-sodium phosphate. A comprehensive review of radiopharmaceuticals is available (31). Specialized pharmacies prepare these products overnight and transport them to hospitals for early administration by members of nuclear medicine departments.

Aerosols. Pressurized containers to deliver aerosolized drug products through appropriate systems of valves and actuators have been available since the 1950s (see AEROSOLS). Such dosage forms are used as external applications of lotions and creams, for oral inhalation, or for treatment of the vaginal cavity, eg, contraceptive foams. Aerosols contain two- or three-phase systems, wherein a volatile liquid or admixture of liquids is sealed in a

container in equilibrium with a vapor phase (propellant). The latter develops pressure to force the liquid from the container through a precisely designed valve upon actuation. If the drug is soluble in the propellant, the system is two-phase. Upon actuation and delivery of the product, the propellant evaporates quickly, and fine dispersion of the drug settles on the area of application. For aerosol products that need accurate dosing, metered valves are used with the valve chamber being recharged between each actuation or dose.

If the drug is not soluble in the propellant, it is dissolved or dispersed in a liquid vehicle. The propellant then constitutes the third phase of the system, and the container must be shaken before valve actuation. Emulsified aerosol products like lotions and creams are examples of such systems.

Aerosols are generally filled cold with chilled product and propellant. This reduces the vapor pressure of the propellant, allowing filling of appropriate containers by volume. The valves are then crimped to the container with the valve in place. The container is then charged with the propellant (under pressure) through the valve. The finished containers are checked for leaks.

The popularity of aerosols has been declining. A widely used group of propellants, the fluorinated hydrocarbons, have been restricted in use since it was found that they can harm the environment by reducing the ozone layer of the upper atmosphere (see AIR POLLUTION; ATMOSPHERIC MODELING; OZONE).

Biotechnology and Dosage Forms. In drug development, biotechnology (qv) generally is recognized as a term that identifies those technologies that utilize living organisms in the production and or alteration of chemical entities that have potential therapeutic activity (32). Besides the production of pharmacologically or biochemically active moieties, these technologies also have been used to produce food ingredients, vaccines, diagnostic testing reagents, and agricultural products (see FERMENTATION; MEDICAL DIAGNOSTIC REAGENTS; VACCINE TECHNOLOGY).

Recombinant deoxyribonucleic (DNA) technology, gene-splicing, genetic engineering (qv), and research in molecular biology and immunology have contributed to biotechnology (33). To track such products that fall into its preview, the FDA considers biotechnology to also include direct DNA transfer technology, hybridoma procedures, cell fusion, molecular alteration of cellular receptors, and the application of cells, cellular components, and tissues that have had their biological activity altered by such technologies (see PROTEIN ENGINEERING; TISSUE ENGINEERING (SUPPLEMENT)) (34).

Drugs developed in the biotechnology arena are peptides and proteins. Erythropoietin, human insulin, and interferons are examples. Generally these chemical entities are present in very small quantities in living organisms or are modifications of such entities. The proteins are produced as solution or injection dosage forms or lyophilized powders to be reconstituted using appropriate vehicles before use. Several specific drugs that have been developed and FDA-approved to date (1995) include human growth factor, human insulin, anticoagulant and thrombolytic agents to treat untoward blood clotting, erythropoietin to treat red blood cell deficiencies, colony stimulating factors to increase white blood cell counts, hepatitis B (HBV) recombinant vaccine and interferons to treat chronic granulomatous diseases, genital warts, hairy cell leukemia, and AIDS-related Kaposi's sarcoma (35,36).

Biotechnology drugs generally are expensive. Erythropoietin, for example, has been estimated to cost approximately \$10,000 yr per patient on renal dialysis. Thus, the cost of erythropoietin-replacement therapy for the estimated 120,000 patients (as of 1995) undergoing hemodialysis could reach \$1 billion (37). Biotechnology also requires large investments, and the patient population needing such drug therapy generally is small which increases the per-dose cost.

Packaging. The packaging components of pharmaceutical products are vital to their safe and effective use. Besides serving the patient as a convenient unit of use, the composite package (unit container, labeling, and shipping components) must provide appropriate identification and necessary information for proper use (including warnings and cautions) and preservation of the product's chemical and physical integrity (see PACKAGING, COSMETICS AND PHARMACEUTICALS).

Labeling. Labeling, controlled by FDA regulations, includes not only the affixed labels, but also the package inserts that provide more detailed information. Trade, generic, or common name, dose, number of dose units present, and name and address of manufacturer and distributor are required. For nonprescription products, adequate directions for use are required. Prescription products must bear the phrase, "Caution: Federal law prohibits use without a prescription" on their labels.

All drug labels must include batch or lot numbers. Using such coded information, products can be traced through all stages of manufacture. Furthermore, the nature of the drug product may require special cautionary phrases, eg, "store in cool place or refrigerator," "protect from light," and "shake well before using." In the 1990s, labels also carry the expiration date, ie, shelf-life. This information is expected to become mandatory.

Labeling information also includes warnings as to possible side effects, eg, drowsiness, and potential harm if used with other drugs or certain foods (drug–drug or drug–food interactions). Inserts are generally intended for use by physicians or pharmacists and give name and description of the product, mode of administration, dosage regimen, therapeutic indications and contraindications, precautions and side effects, units of supply, and literature citations. All labeling must be approved by the FDA as part of the New Drug Application. The FDA has proposed the widespread use of patient package inserts (PPIs). These are separate sheets of information, written in layman's terms, providing more detailed information. To date (1995), however, they have not been mandated.

Containers. The USPXXIII–NFXVIII lists container requirements such as well-closed, tight, or light-resistant. Most containers are light-resistant (amber) glass or plastic. The latter is break-resistant and lightweight, which reduces shipping costs and increases safety.

In hospitals and long-term care units, unit-dose packages are used more and more. This system allows better control of the dispensed drugs in institutional settings and precludes the dispensing of larger numbers of doses than needed.

Quality Control and Quality Assurance

Quality control (QC) involves the regular, daily assessment and or analysis, according to established protocols and standards, of all ingredients, processes, and finished products (38,39). Official USP NF monographs, for example, provide various chemical, physical, and biological tests and specifications for assurance of purity, potency, and stability of component ingredients used to prepare and package drug products. In-process testing performed at specified points during the production stages, eg, tablet disintegration, weight, fragility, and hardness tests, are part of the QC program. Likewise, the FDA requires process validation procedures as QC constituents. The FDA also monitors QC standards through the requirements of the Current Good Manufacturing Procedures regulations.

Quality assurance (QA) constitutes the broad, oversight functions that include the auditing of the various QC functions. This ensures that appropriate QC standards have been developed and are in regular use (38,39). It includes the acknowledgment of necessary change, when needed, to maintain quality, the education of personnel to be sure they know their assigned QC responsibility, the documentation of all such training programs, spot auditing as an assurance process, and appropriate action to assure and maintain compliance with all externally imposed regulations and internally established criteria and standards.

BIBLIOGRAPHY

"Pharmaceuticals" in *ECT* 1st ed., Vol. 10, pp. 228–263, by H. M. Merker and A. R. Whale, Parke, Davis & Co.; in *ECT* 2nd ed., Vol. 15, pp. 112–132, by J. J. Windheuser and J. Perrin, The University of Wisconsin; in *ECT* 3rd ed., Vol. 17, pp. 272–290, by P. Zanowiak, Temple University.

1. *Trends in U.S. Pharmaceutical Sales and R & D: 1990–93 PMA Annual Survey Report*, Pharmaceutical Manufacturers Association, Washington, D.C., 1993.
2. *The U.S. Generic Drug Industry: Brand-Name and Generic U.S. Sales*, Generic Pharmaceutical Industry Association, Washington, D.C., 1992.
3. *Percent of Total U.S. Prescriptions Dispensed Generically: 1985–97*, Generic Pharmaceutical Industry Association, Washington, D.C., 1992.
4. *OTC Retail Sales (1964–1994)*, Nonprescription Drug Manufacturers Association, Washington, D.C., 1995.
5. *Retail Perspective and Provider Perspective*, IMS America, Ltd., Plymouth Meeting, Pa., 1993.
6. *New Drug Approvals in 1994*, Pharmaceutical Research and Manufacturers of America, Washington, D.C., 1995.
7. *U.S. Public Law 98-417*, 98th Congress, 98 Stat. 1585, U.S. Government Printing Office, Washington, D.C., Sept. 24, 1984.
8. M. Freudenheim, "Big Drug Companies Brace for Battles on Patent Rights," *New York Times*, New York, June 8, 1995, p. D2.

9. "FDA Gatt Policy Prevents ANDA Approvals During Extended Patent Period," *FDC Reports*, Washington, D.C., May 29, 1995, pp. 3–5.

10. *OTC Facts and Figures*, Nonprescription Drug Manufacturers Association, Washington, D.C., 1995.

11. W. E. Gilbertson, in T. R. Covington, ed., *Handbook of Nonprescription Drugs*, 10th ed., American Pharmaceutical Association, Washington, D.C., 1993, pp. 21–37.

12. H. C. Ansel and N. G. Popovich, *Pharmaceutical Dosage Forms and Drug Delivery Systems*, 5th ed., Lea & Febiger, Philadelphia, Pa., 1990, pp. 92–133.

13. G. S. Banker, in G. S. Banker and C. T. Rhodes, eds., *Modern Pharmaceutics*, 2nd ed., Marcel Dekker, New York, 1990, pp. 15–20.

14. P. Zanowiak, in J. Swarbrick and J. C. Boylon, eds., *Encyclopedia of Pharmaceutical Technology*, Vol. 4, Marcel Dekker, New York, 1991, pp. 209–229.

15. P. Zanowiak, in *Ullmann's Encyclopedia of Industrial Chemistry*, VA19, VCH Verlagsgesellschaft, MbH, D-6940 Weinheim, Germany, 1991, pp. 241–271.

16. Ref. 12, pp. 96–99.

17. W. E. Gilbertson, in E. G. Feldmann and W. L. Blockstein, eds., *Handbook of Nonprescription Drugs*, 9th ed., American Pharmaceutical Association, Washington, D.C., 1990, pp. 25–39.

18. *Fed. Reg.* **45**, 18398 (Mar. 1980).

19. J. S. Welsh, *Am. Pharm.* **NS20**(9), 11 (1981).

20. *Approved Prescription Drug Products, with Therapeutic Equivalence Evaluations*, Bureau of Drugs, Food and Drug Administration, U.S. Dept. of Health and Human Services, Bethesda, Md., 1980.

21. V. V. Ranade, *J. Clin. Pharmacol.* **31**, 98–115 (1991).

22. Ref. 12, pp. 134–196.

23. "Pharmaceutical Dosage Forms (1151)", in *United States Pharmacopeia XXIII and National Formulary XVIII*, U.S. Pharmacopeial Convention, Washington, D.C., 1995, pp. 1939–1951.

24. R. A. Buerki and G. J. Higby, in Ref. 14, Vol. 7, 1993, pp. 299–339.

25. K. Heilmann, *Therapeutic Systems: Rate-Controlled Drug Delivery, Concept and Development*, 2nd rev. ed., Thieme-Stratton, Inc., New York, 1984.

26. K. E. Avis, in A. R. Gennaro, ed., *Remington: The Science and Practice of Pharmacy Sciences*, Vol. 2, 19th ed., Mack Publishing Co., Easton, Pa., 1995, pp. 1524–1548.

27. *USP/NF*, 23rd and 18th eds., U.S. Pharmacopeial Convention, Washington, D.C., 1995, pp. 1650–1652.

28. S. J. Turco, in Ref. 14, pp. 231–247.

29. S. J. Turco, ed., *Sterile Dosage Forms: Their Preparation and Clinical Application*, 4th ed., Lea & Febiger, Philadelphia, Pa., 1994, pp. 1–25.

30. S. J. Turco, ed., in Ref. 29, pp. 344–353.

31. E. D. Mackowiak, in Ref. 29, pp. 299–315.

32. Ref. 12, pp. 299–302.

33. A. H. der Marderosian, in Ref. 26, Vol. 1, pp. 809–828.

34. H. I. Miller, *Biotechnology Medicines in Development*, a 1991 Survey Report, Pharmaceutical Manufacturers Association, Washington, D.C., 1991, p. 12.

35. S. J. Turco, in Ref. 26, pp. 1549–1562.

36. J. D. Leary, in L. Shargel, ed., *Comprehensive Pharmacy Review*, 2nd ed., Harwal Publishing, Philadelphia, Pa., 1994, pp. 177–182.

37. R. F. Doolittle, *N. Eng. J. Med.* **324**(19), 1360–1361 (1991).

38. C. R. Erskine and K. A. Herzog, in Ref. 33, pp. 648–652.

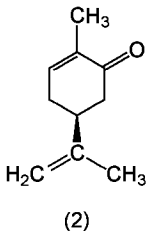
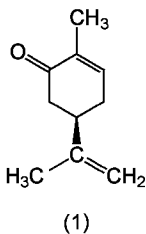
39. E. R. Anderson, in Ref. 36, p. 406.

Paul Zanowiak
Temple University

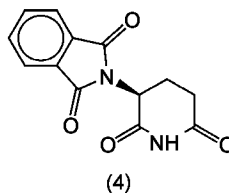
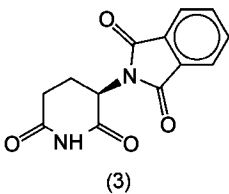
PHARMACEUTICALS, CHIRAL

Stereospecific structure-activity relationships are of sufficient complexity to warrant a clear understanding of the terms involved. Prior to discussing these relationships the following definitions are offered. Stereoisomers are compounds which have the same molecular formula but differ in the arrangement of their atoms in space. Chiral compounds are compounds which have nonsuperimposable mirror images. Enantiomers are pairs of stereoisomers which are nonsuperimposable mirror images; they possess identical physical and chemical properties within an achiral environment. Stereoisomers other than enantiomers, ie, diastereomers, are identified by distinct physical and chemical properties including melting points, spectral characteristics, and rates of reaction with both chiral and achiral reactants. Enantiomers, however, are only distinguished when in the presence of a homochiral environment such as polarized light, chiral solvents, chiral reagents, or chiral molecules such as biomolecules, eg, nucleic acids (qv), proteins (qv), and carbohydrates (qv). The two molecules in a pair of enantiomers rotate a plane of polarized light with equal intensities, but in opposite directions. The dextrorotatory isomer (+ or *d*) rotates the plane of polarized light clockwise; the levorotatory isomer (- or *l*) rotates the plane of polarized light counterclockwise. An equal mixture of (+) and (-)-enantiomers is a racemic mixture or racemic compound and does not rotate a plane of polarized light. Optical rotation, an intrinsic property of the substance, has no bearing on drug-macromolecule interactions. It is the absolute configuration of the homochiral compound that is important for its interaction with biomolecules.

Absorption, metabolism, and biological activities of organic compounds are influenced by molecular interactions with asymmetric biomolecules. These interactions, which involve hydrophobic, electrostatic, inductive, dipole-dipole, hydrogen bonding, van der Waals forces, steric hindrance, and inclusion complex formation give rise to enantioselective differentiation (1,2). Within a series of similar structures, substantial differences in biological effects, molecular mechanism of action, distribution, or metabolic events may be observed. For example, (*R*)-carvone [6485-40-1] (1) has the odor of spearmint whereas (*S*)-carvone [2244-16-8] (2) has the odor of caraway (3,4).



The amino acids L-leucine, L-phenylalanine, L-tyrosine, and L-tryptophan all taste bitter, whereas their D-enantiomers taste sweet (5) (see AMINO ACIDS). D-Penicillamine [52-67-5], a chelating agent used to remove heavy metals from the body, is a relatively nontoxic drug effective in the treatment of rheumatoid arthritis, but L-penicillamine [1113-41-3] produces optic atrophy and subsequent blindness (6). L-Penicillamine is roughly eight times more mutagenic than its enantiomer. Such enantioselective mutagenicity is likely due to differences in renal metabolism (7). (*R*)-Thalidomide (3) is a sedative-hypnotic; (*S*)-thalidomide (4) is a teratogen (8).



Unfortunately, (*R*)-thalidomide, containing some of the (*S*)-enantiomer, was given to pregnant women resulting in a large number of fetal deaths and congenital malformations. It is still unclear whether the administration of pure (*R*)-thalidomide (3) would have prevented the observed teratogenicity. The stereocenter of thalidomide is labile and may racemize under physiological conditions (9).

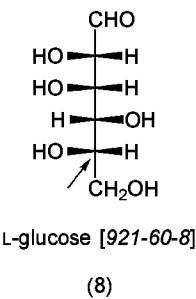
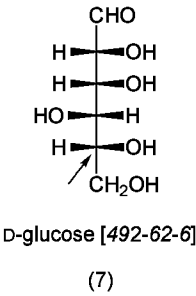
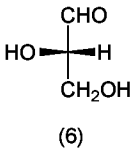
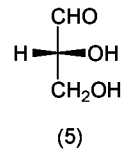
Care should be exercised when attempting to interpret *in vivo* pharmacological data in terms of specific chemical-biological interactions for a series of asymmetric compounds, particularly when this interaction is the only parameter considered in the analysis (10). It is important to recognize that the observed difference in activity between optical antipodes is not simply a result of the association of the compound with an enzyme or receptor target. Enantiomers differ in absorption rates across membranes, especially where active transport mechanisms are involved (11). They bind with different affinities to plasma proteins (12) and undergo alternative metabolic and detoxification processes (13). This ultimately leads to one enantiomer being more available to produce a therapeutic effect.

The importance of optical isomers with regard to biological effect has a long history (14) beginning with the observations of Pasteur, who by hand separated nonsuperimposable mirror image crystals of (+) and (-)-sodium ammonium tartrate (15). Subsequently, Pasteur showed that these enantiomers are effectively differentiated by molds and yeasts. Dramatically improved techniques in asymmetric syntheses (16), chiral separations (17), analytical techniques, and stereochemical characterization have led to the widespread production and biological evaluation of numerous homochiral molecules (see ANALYTICAL METHODS). The appearance in the late 1980s of the journals *Chirality* and *Tetrahedron Asymmetry* provide testament to this vastly expanding research area (18–20). The U.S. FDA requires that both enantiomers of a drug be individually tested when associated toxicities occur near the effective dose of the racemic substance (21,22). It has been suggested that the use of racemic drugs in human subjects cannot be justified until both enantiomers are tested thoroughly, both individually and in composite mixtures. Often, side effects of therapeutics are not discovered until after large-scale marketing (18). The distomer (therapeutically inactive enantiomer) may be at best a nontoxic impurity, but is often associated with dangerous side effects as exemplified by the thalidomide problem.

The thrust toward homochiral drugs by leading researchers and organizations such as the FDA, the rapidly expanding technology of asymmetric syntheses and chiral separations, the decreased side effects found with homochiral drugs, and the potential financial benefits are expected to ensure that the majority of chiral synthetic drugs will, in the future, be available in enantiomerically pure form (see PHARMACEUTICALS; RESEARCH TECHNOLOGY MANAGEMENT).

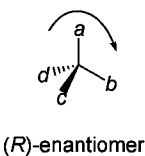
Background

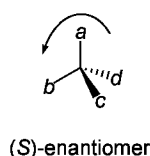
Nomenclature. Compounds which have tetrahedral atoms having four different substituents are often chiral. These tetrahedral atoms are referred to as stereocenters or stereogenic atoms; the terms asymmetric atom or asymmetric center are considered misnomers. The letters D and L are used to denote the absolute configurations of amino acids and sugars according to Fischer-Rosanoff nomenclature (qv) (23) (see SUGAR). In this system, dextrorotatory glyceraldehyde [453-17-8] (5) is arbitrarily assigned an absolute configuration of D (24). In a Fischer projection, the most highly oxidized carbon is placed on top and the last stereocenter determines the absolute configuration L or D. Examples of Fischer projections are shown for D- and L-glyceraldehyde (6) and D-glucose [492-62-6] (7) and L-glucose [921-60-8] (8) where the arrow denotes the determining stereocenter.



If the heteroatom attached to the last stereocenter projects to the right, the compound is of the D-configuration; if the heteroatom points to the left, the compound is of the L-configuration (23). There is no simple relationship between sign of rotation (*d*(+) or *l*(-)) and the absolute configuration D or L. However, optical activity may be related empirically to absolute configuration by observing changes in optical rotation with varying wavelength, ie, optical rotatory dispersion (ord) and circular dichroism (cd) (25).

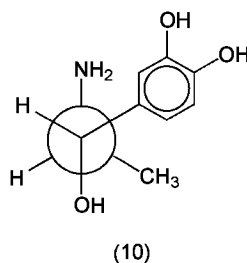
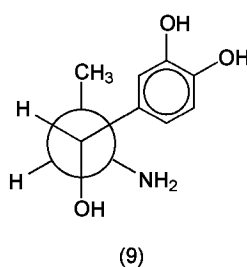
Because of ambiguities involved in this nomenclature, the Cahn-Ingold-Prelog rules were introduced and are widely used to designate the absolute configuration of stereocenters (26). Groups attached to the stereogenic atom are assigned priorities according to the atomic number of the atom attached to the stereogenic center. Highest priority is given to the atom with the highest atomic number. The molecule is drawn such that the function of lowest priority (*d*) is directed away from the viewer:



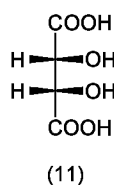


If the observed order of priority of the remaining three functions ($a > b > c$) is in a clockwise direction, the absolute configuration is designated R (rectus or right); if counterclockwise, the configuration is S (sinister or left). The concepts of stereochemistry and chirality have been extensively discussed and reviewed (27–29).

For compounds with n stereocenters, 2^n stereoisomers are possible. These stereoisomers consist of a set of diastereomers and the enantiomer of each diastereomer. Diastereomers with two stereocenters may be named using erythro and threo terminologies. Erythro indicates that the highest priority groups are on the same side of the molecule or in a cis relationship when the lowest priority groups of both stereocenters are directed away from the viewer, eg, (1*S*,2*R*)-2-methylnoradrenaline (9). Threo indicates that the highest priority groups are on opposite sides of the molecule or in a trans relationship, eg, (1*S*,2*S*)-2-methylnoradrenaline (10).



When vicinal, ie, adjacent, stereogenic carbons have identical functional groups in an erythro relationship, the term meso is used as in *meso*-tartaric acid (11).



Diastereomers which differ at a single stereocenter are called epimers.

Enantiomeric purity, measured as the enantiomeric excess (ee) of an isomer, is determined by the formula ($^\circ$ of major isomer) - ($^\circ$ of minor isomer). Thus, if a chiral drug is said to be of 50% ee, the composite mixture contains 75% of one enantiomer and 25% of the other. Enantioselectivity refers to the greater activity of one enantiomer over its mirror image. Enantiospecificity is rarely observed and implies that one enantiomer possesses 100% of the observed activity (30); in most cases it is more accurate to use the term highly enantioselective. The pharmacologically more active enantiomer is termed the eutomer and the less active enantiomer is referred to as the distomer. The eudismic ratio is the ratio of the potency of the eutomer versus the potency of the distomer for a specific biological or pharmacological action. Generally, highly potent drugs possess high eudismic ratios, a phenomenon referred to as Pfeiffer's rule (19). The eudismic proportion is the ratio of the concentrations of eutomer and distomer ([eutomer] [distomer]) within a specific tissue or fluid. This is a property determined by enantioselective plasma protein binding, active transport mechanisms, and metabolism (31).

The therapeutic efficacy of a drug is generally measured in terms of ED₅₀ or ID₅₀ which represent the concentration of drug which produces 50% of the maximum effect or 50% of maximum inhibition. LD₅₀ represents the concentration of drug that produces 50% fatalities in test animals. The therapeutic index is the ratio of the ED₅₀ versus LD₅₀. Detailed descriptions of the terminology and fundamental principles of pharmacology are available (32) (see PHARMACODYNAMICS).

Role of Homochiral Molecular Building Blocks. Generally L-amino acids and D-sugars are found in biological systems (33,34). The evolution of stereochemical preferences in biological processes raises several interesting questions. What is the origin of this biochemical stereoselectivity (35), ie, why are proteins built from L-amino acids and not D-amino acids? What is the significance of homochiral proteins and carbohydrates? What important biological roles do D-amino acids and L-sugars possess? Quantum mechanics and mathematical calculations demonstrate that L-amino acids have a slightly lower energy than D-amino acids (34,36). Similarly D-glyceraldehyde (5), which serves as a precursor for other sugars, is at a lower energy than the corresponding L-glyceraldehyde (6). These energy differences, estimated to be about 10^{-14} J/mol, arise from the influence of subatomic particles, ie, bosons, quarks, neutrinos, electrons, etc (36,37). This energy difference, however slight, may have favored the enantioselective formation of L-amino acids and D-sugars and their subsequent selective incorporation into biomolecules.

The formation of D-amino acids in polypeptides and in monomeric form during processing of proteinaceous foods has raised considerable concern about associated nutritional and toxicity effects (37). Racemization of amino acids occurs under strongly basic or acidic conditions, which are conditions used in food processing (qv). The presence of aldehydic contaminants enhances the rate of amino acid racemization through formation of stereogenically labile α -imino acids (12) and (13) (Fig. 1). D-Amino acids may be utilized in a nutritional manner if they are converted to L-amino acids. Racemase enzymes, observed only in microorganisms, interconvert L- and D-amino acids. Mammals must rely on D-amino acid oxidases to catalyze D- to L-amino acid conversions. Oxidase-catalyzed deamination to α -keto acids (16) and subsequent stereoselective amination produces the nutritionally valuable L-amino acids (38). Peptides containing D-amino acid residues are considerably less prone to peptidase hydrolysis than the corresponding all-L-peptides, and this results in the excretion of the peptides without nutritional benefit (38). For the most part, D-amino acids are generally no more toxic than their L-enantiomers. Thus, foods containing proteins with high concentrations of D-amino acid residues may be useful for weight management (38).

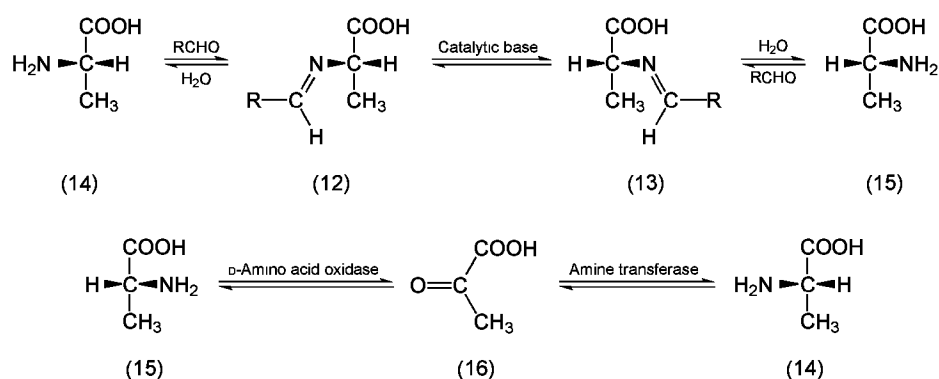
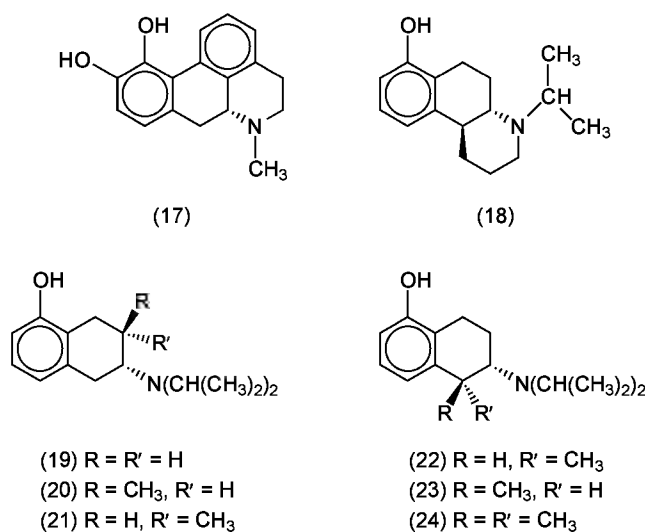


Fig. 1. Processes for the interconversion of D- and L-amino acids.

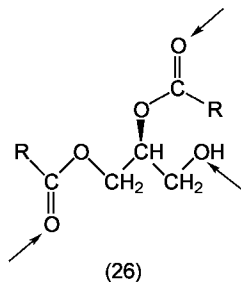
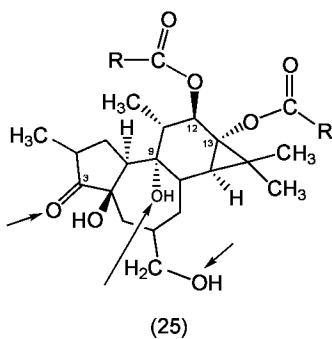
Modeling of Drug-Receptor Interactions. The identification of molecular interactions between drugs and their receptor or enzyme targets and the three-dimensional spatial requirements of the macromolecular binding pocket are important for the rational design of new, more selective, and potent pharmaceuticals (39). Methods used to explore such interactions include nmr spectroscopy of receptor–ligand complexes, molecular modeling, point mutation analysis, and binding assays of conformationally constrained, stereochemically defined small molecules (40) (see MAGNETIC SPIN RESONANCE; MOLECULAR MODELING (SUPPLEMENT)). X-ray crystal structure analysis of macromolecule–ligand complexes provides information concerning important molecular interactions which give rise to the observed affinity between the macromolecule and ligand. Computers are used to graphically display calculated crystal structures. Numerous computer programs have been developed and refined which are capable of determining the energy minimized structures of such complexes. These programs, which take into account numerous variables such as bond lengths, bond angles, hydrogen bonding, steric influences, dipole–dipole interactions, etc, provide a basis for studying and predicting the affinities new drugs will possess for their macromolecular targets (41). By using such technology it is often possible to visually discern why enantiomers bind with different affinities to their macromolecular targets.

Molecular modeling techniques, although aesthetically pleasing, are far from reliable owing to the large number of associated variables. Results from molecular modeling studies may be confirmed by comparing calculated molecular binding energies with experimentally obtained binding constants for a series of high affinity compounds. The active sites of receptors and enzymes which have not yet been crystallized may be reliably mapped by such high affinity probes. Furthermore, molecular modeling of known high affinity ligands is also useful in determining potential pharmacophore binding conformations of the substrate, ie, the conformation of functional groups on the substrate which interact with the macromolecular species (enzymes or receptors).

Molecular modeling has provided information about the binding conformation of dopamine D₂-receptor agonists. Such information is useful for the rational design of new highly selective dopamine D₂-receptor agonists. The relative positions of the phenolic hydroxyl, aromatic ring, and amine functionalities, required for binding of these receptor agonists, were determined by the active analogue approach (42) (Fig. 2). For this analysis roughly 20 electronically (MMP2 software) calculated energy minimized conformations were determined for the potent agonists (R)-apomorphine [58-00-4] (17) and the tricyclic amine (18). Only one low energy conformer existed, however, in which the two agonists were superimposable. This conformer was rationalized to have the conformation needed for binding. Subsequent studies provide evidence substantiating this conclusion. A series of 5- and 7-substituted 6-(N,N-diisopropyl)amino-1-(5,6,7,8-tetrahydro)naphthalenol derivatives (19–24) were synthesized. Those compounds capable of adopting the predicted binding conformation most easily, ie, had the lowest energy requirements for this conformation, were the most potent, ie, (19), (21), and (22) (42).

Fig. 2. Molecular modeling of dopamine D₂-receptor agonists used to define the molecular conformation needed for selective high affinity binding.

The spatial and steric requirements for high affinity binding to protein kinase C (PKC), a macromolecule that has not yet been crystallized, were determined. Protein kinase C plays a critical role in cellular signal transduction and is in part responsible for cell differentiation. PKC was identified as the macromolecular target for the potent tumor-promoting phorbol esters (25). The natural agonists for PKC are diacylglycerols (DAG) (26). The arrows denote possible sites of interaction.

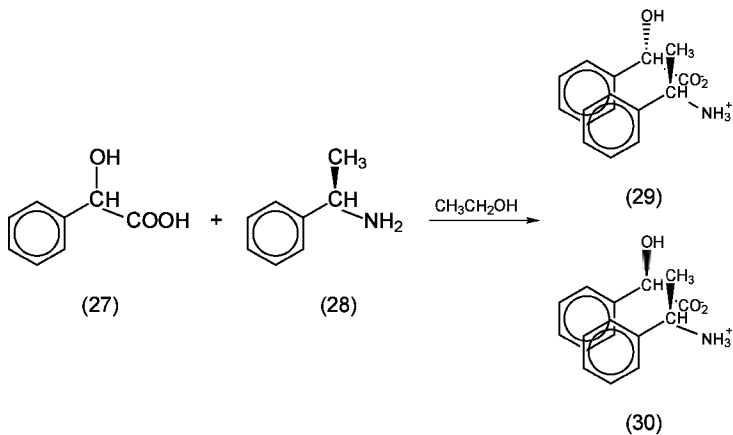


Molecular modeling demonstrates that phorbol esters act as rigid DAG analogues, in which the pharmacophore is made up of the C₃ carbonyl, and the C₉ and C₂₀ hydroxyl functionalities. Energy minimization of the x-ray crystal structure of phorbol 12,13-dibutyrate provides a basis on which rigid analogues have been designed and synthesized. The relative binding affinities of these analogues were determined (43), and a correlation between the calculated energy of binding and experimentally determined binding affinities was observed. This work supports the proposed binding conformation of phorbol 12,13-dibutyrate and provides a basis for the design of simplified, novel PKC agonists.

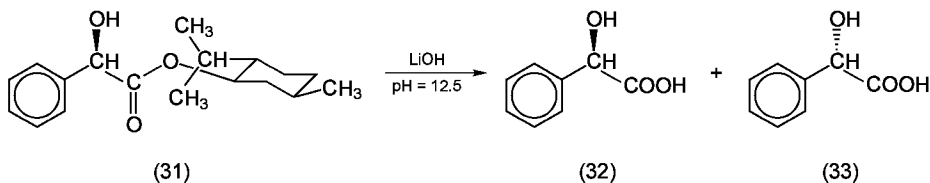
Molecular modeling demonstrates that those compounds which most easily adopt the proposed binding conformation possess the greatest affinity for their specific macromolecular target. Similarly, enantiomers bind to the enzyme or receptor targets in a preferred conformation. Often one enantiomer of the pair requires less energy to adopt the preferred binding conformation and accordingly shows greater affinity for its macromolecular target.

Methods for the Preparation of Homochiral Drugs

Resolution Methods. Chiral pharmaceuticals of high enantiomeric purity may be produced by resolution methodologies, asymmetric synthesis, or the use of commercially available optically pure starting materials (44,45). Resolution refers to the separation of a racemic mixture. Classical resolutions involve the construction of a diastereomer by reaction of the racemic substrate with an enantiomerically pure compound. The two diastereomers formed possess different physical properties and may be separated by crystallization (qv), chromatography (qv), or distillation (qv). A disadvantage of the use of resolutions is that the best yield obtainable is 50%, which is rarely approached. However, the yield may be improved by repeated racemization of the undesired enantiomer and subsequent resolution of the racemate. Resolutions are commonly used in industrial preparations of homochiral compounds (16). Chiral acids and amines are generally separable by crystallization of the diastereomeric salts formed with an appropriate optically pure amine or acid, respectively (46). Racemic mixtures of mandelic acid (27) are resolved by treatment with optically pure (R)-(+)-methylbenzylamine (28) and formation of the diastereomeric salts (29) and (30). (R)-(-)-Mandelic acid selectively crystallizes with (R)-(+)-methylbenzylamine (28) and is isolated by simple filtration.

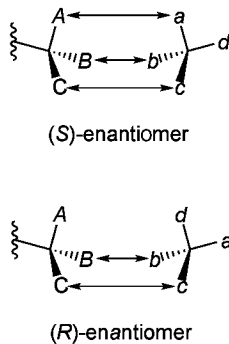


Three general methods exist for the resolution of enantiomers by liquid chromatography (qv) (47,48). Conversion of the enantiomers to diastereomers and subsequent column chromatography on an achiral stationary phase with an achiral eluant represents a classical method of resolution (49). Diastereomeric derivatization is problematic in that conversion back to the desired enantiomers can result in partial racemization. For example, (1R,2S,5R)-menthol (R)-mandelate (31) is readily separated from its diastereomer but ester hydrolysis under numerous reaction conditions produces (R)-(-)-mandelic acid (32) which is contaminated with (S)-(+)-mandelic acid (33).



Direct resolution of the enantiomers without derivatization is performed by use of an achiral stationary phase with a chiral mobile phase or, more commonly, by use of a chiral stationary phase and an achiral mobile phase. Ligand-exchange chromatography (50–52), using Cu²⁺ proline, and chiral ion pair chromatography, which involves use of a chiral counterion in the mobile phase, exemplify the former method. Chromatographic resolution of enantiomers using chiral stationary phases is advantageous over the previously described methods in that no derivatization is required, there is no need to

employ expensive mobile phases, and the method does not require complicated product analysis. Stationary phases include cyclodextrins, protein bonded supports, chiral polymers, and the Pirkle type (53–56) (see POLYMERS). Pirkle columns involve the attachment of a low molecular weight chiral molecule to a solid support. Diastereomeric interactions between the racemate and the column alter the elution times between the two enantiomers making separation possible. The greater association of one enantiomer than the other with the chiral stationary phase allows for enantiomeric differentiation and separation. The three-point rule for chiral recognition is used to rationalize the separation of enantiomers by use of a homochiral stationary phase and an achiral eluent. The (*S*)-enantiomer possesses three favorable interactions with the stationary phase and is therefore expected to traverse the column at a slower rate than its (*R*)-enantiomer, which maintains only two favorable interactions (53).



Enzymes are used in organic syntheses because they catalyze reactions under gentle conditions, ie, ambient temperatures and pressure and neutral pH, with high enantioselectivity; in addition, they are inexpensive and can be recycled (44,57,58) (see ENZYME APPLICATIONS, INDUSTRIAL; ENZYMES IN ORGANIC SYNTHESIS; MICROBIAL TRANSFORMATIONS). The discovery that enzymes catalyze reactions in organic solvents has greatly enhanced the potential utility of these chiral catalysts (59). Hydrolase enzymes such as pig liver esterase (PLE) show broad substrate specificity and are particularly useful for labile substrates; they are used for the enantioselective hydrolysis of prochiral diesters and the kinetic resolution of racemates via ester hydrolysis (60). PLE-catalyzed hydrolysis of dimethyl *cis*-cyclohex-4-ene-1,2-dicarboxylate [4841-84-3] (34) gives the mono-acid methyl (1*R*,2*S*)-2-carbomethoxycyclohex-4-ene-1-carboxylic acid [88335-94-8] (35) in nearly 95% yield and 90% ee (Fig. 3). Treatment of racemic methyl 3,4-epoxybutanoate [4509-09-5] (36) with PLE yields (*S*)-3,4-epoxybutanoic acid [109462-43-3] (37) and methyl (*R*)-3,4-epoxybutanoate [109462-42-2] (38). Enantioselective esterifications can also be carried out enzymatically. (*S*)-2-(2,2-Diethoxyethane)-1-acetyloxy-3-propanol [134665-24-0] (40), a useful synthon for the production of homochiral three-substituted butanolides, is prepared by transesterification of 2-(2,2-diethoxyethane)-1,3-propanediol [55387-85-4] (39) (61).

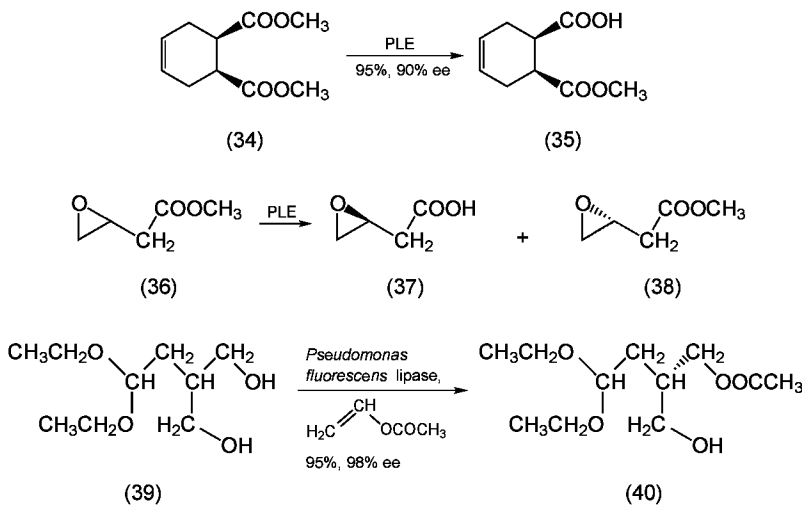
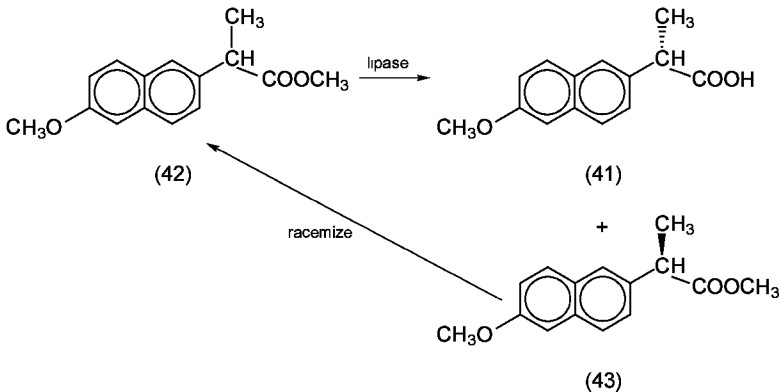
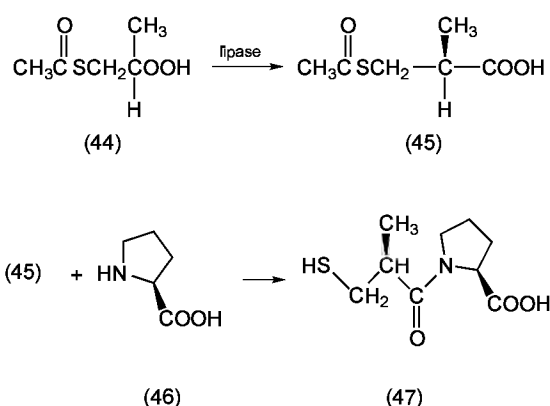


Fig. 3. Enzyme-catalyzed resolutions. PLE = pig liver esterase. See text.

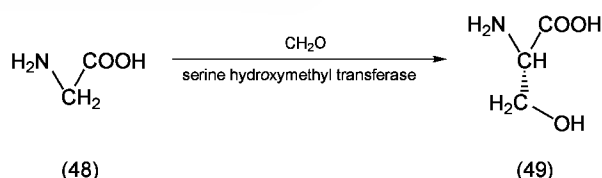
Numerous synthetic transformations have been carried out using enzymes. Oxidoreductases are commonly used to reduce prochiral ketones with high enantioselectivity. The addition of hydrogen cyanide to aldehydes, catalyzed by the enzyme mandelonitrile lyase, yields the corresponding (*R*)-cyanohydrins. Nitrile hydrolase converts cyano functionalities into the corresponding carboxylic acids, a transformation which usually requires harsh conditions, such as treatment with concentrated hydrochloric acid for 24 hours (62). A principal disadvantage in the use of enzymes is that they produce only one enantiomer; it is not always straightforward to produce the other optical isomer. However, in industry this is not problematic because large-scale production of only one isomer is usually required (63). Naproxen [22204-53-1] (41), a nonsteroidal antiinflammatory drug marketed as the pure (*S*)-enantiomer, is produced from the methyl ester (42) in better than 98% ee using the enzyme *Candida cylindracea* lipase (64). The (*R*)-ester (43) is subsequently racemized and retreated with the enzyme to optimize the yield of the (*S*)-enantiomer.



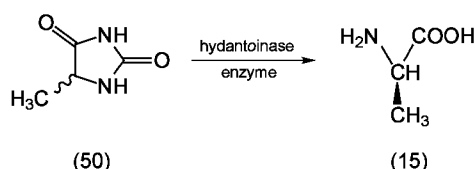
One of the homochiral starting materials (45) for the acetylcholinesterase (ACE) inhibitor captopril [62571-86-2] (47) is produced by a lipase enzyme-catalyzed resolution of racemic 3-methyl-4-acetylthiobutyric acid (44) and *S*-proline (46) (65).



Several strategies for the production of pure D- or L-amino acids rely on the use of enzymes. L-Serine (49) is synthesized by combining glycine (48) and formaldehyde in the presence of the enzyme serine hydroxymethyl transferase (66).

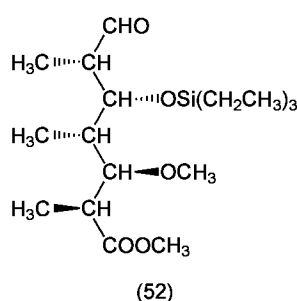
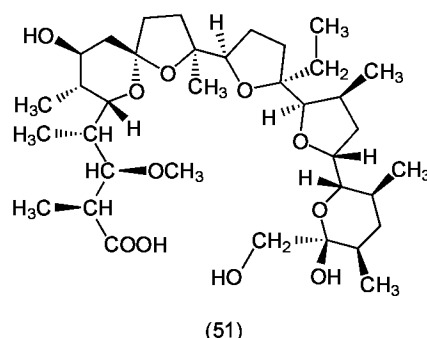


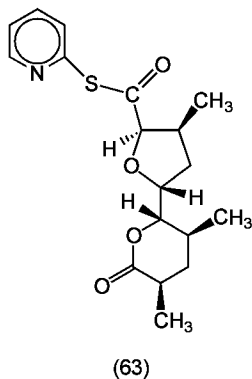
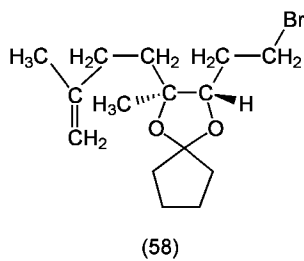
Both pure L- and D-amino acids can be made using hydantoinase enzymes. These enzymes catalyze the stereoselective hydrolysis of racemic hydantoins such as (50) which is used for the production of D-alanine (15) (58).



The high degree of stereoselectivity observed with enzyme reactions provides further evidence as to the importance of drug stereochemistry for pharmaceutical activity.

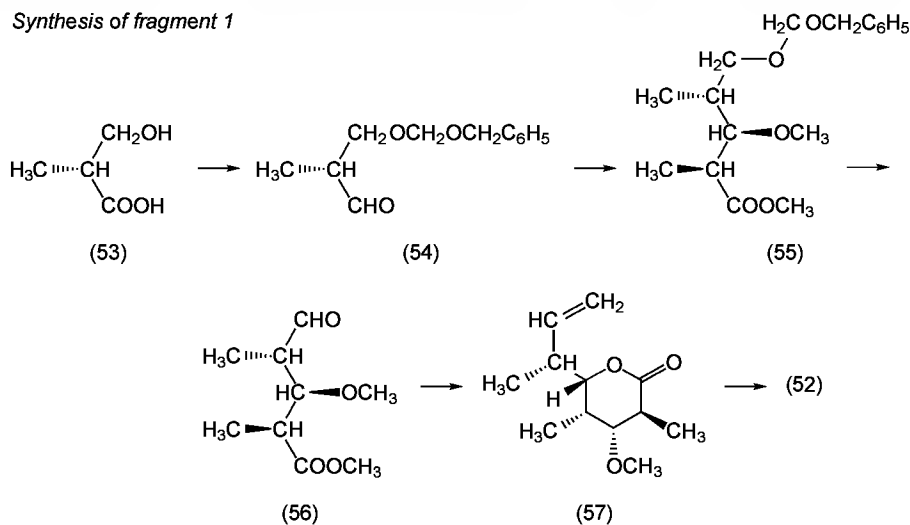
Methods Employing Enantiomerically Pure Starting Materials. A large number of optically pure natural products are commercially available and relatively inexpensive. Amino acids, carbohydrates, and terpenes are some of the homochiral building blocks used routinely in enantiomeric syntheses (67–69). The stereocenter(s) in such building blocks are used as either chiral synthons (chirons) or as chiral auxiliaries. Chirons are incorporated directly into the desired target molecule; chiral auxiliaries are used as an external handle to direct the formation of other stereocenters within the desired molecule and are subsequently cleaved. Chirons are an inexpensive source of chirality and their use in organic synthesis generally produces enantiomerically pure compounds of known absolute configuration. The commercially available polyether antibiotic monensin [17090-79-8] (51) (Fig. 4), which contains 17 stereocenters, is synthesized by the preparation and coupling of three fragments, (52), (58), and (63), each prepared from commercially available optically active starting materials (70,71).





The starting material for synthesis of fragment 1 (52) is (+)-β-hydroxyisobutyric acid [1910-47-0] (53), which is protected and reduced to the aldehyde (54). Diastereo-selective aldol reaction with 2-methyl-2-triethylsilyloxy-3-pentanone, oxidative cleavage, and dimethylation produces the optically pure diastereomer (55) in 50% yield. Deprotection and oxidation produces aldehyde (56), which undergoes a second aldol reaction with *cis*-2-butenyldiethylaluminum at -78°C followed by spontaneous lactonization to yield lactone intermediate (57). This intermediate contains five stereocenters of the desired configuration for the production of monensin. Lactone (57) is converted to fragment 1 by ozonolysis of the corresponding methyl ester. Fragment 2, the spiroketal (58), is prepared from natural (*S*)-(-)-malic acid [97-67-6] (59). Regioselective acetonide formation, followed by reduction of the free carboxylic acid, acid-catalyzed lactonization, and α-hydroxyl group protection produces the homochiral lactone (60). Grignard reaction with methyl magnesium bromide, and primary alcohol protection with *tert*-butyldimethylsilylchloride (TBS) yields ketone (61). Highly diastereoselective addition of 3-methyl-3-butenylmagnesium bromide affords the threo adduct (62) in 96% ee. This compound is subsequently converted to fragment 2. Delta-lactone (63), fragment 3, is synthesized in a convergent manner by coupling aldehyde (64), prepared from (*R*)-citronellic acid [18951-85-4] (65), with the chiral triphenylphosphine derivative (66), synthesized from (*R*)-β-hydroxyisobutyraldehyde [38433-80-6] (67). Wittig reaction produces the *cis*-olefin (68), which is subsequently converted to fragment 3. Fragments 2 (58) and 3 (63) are combined via a Grignard reaction which yields ketone (69). Further manipulation yields intermediate (70). Aldol reaction of ketone (70) with fragment 1 provides polyol (71) which is converted to monensin (51) by deprotective hydrogenation, acid-catalyzed spiroketalization, and ester saponification (72). The synthesis of monensin provides an outstanding example of how chirons are exploited and manipulated for the construction of stereogenically complex molecules.

Synthesis of fragment 1



Synthesis of fragment 2

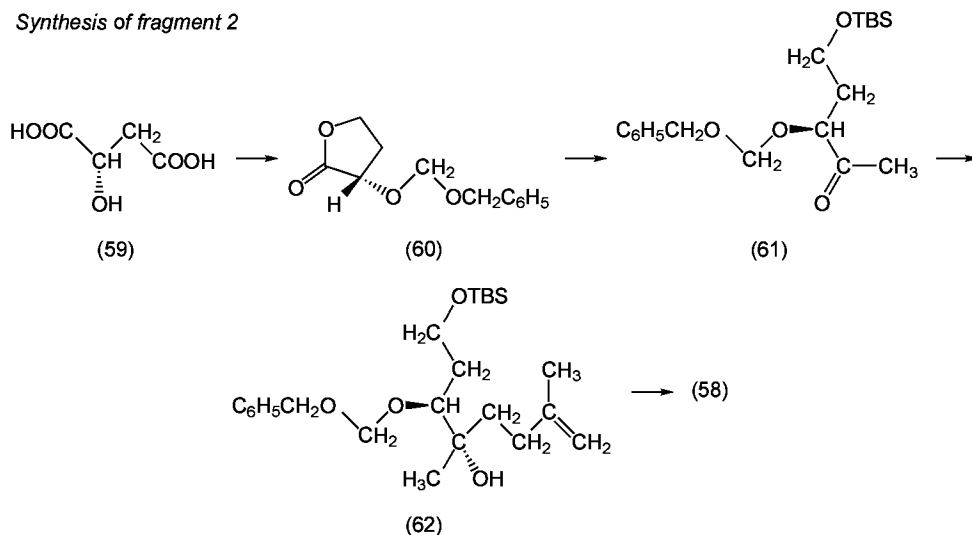


Fig. 4. Synthesis of monensin from homochiral natural products; TBS *tert*-butyldimethylsilyl chloride.

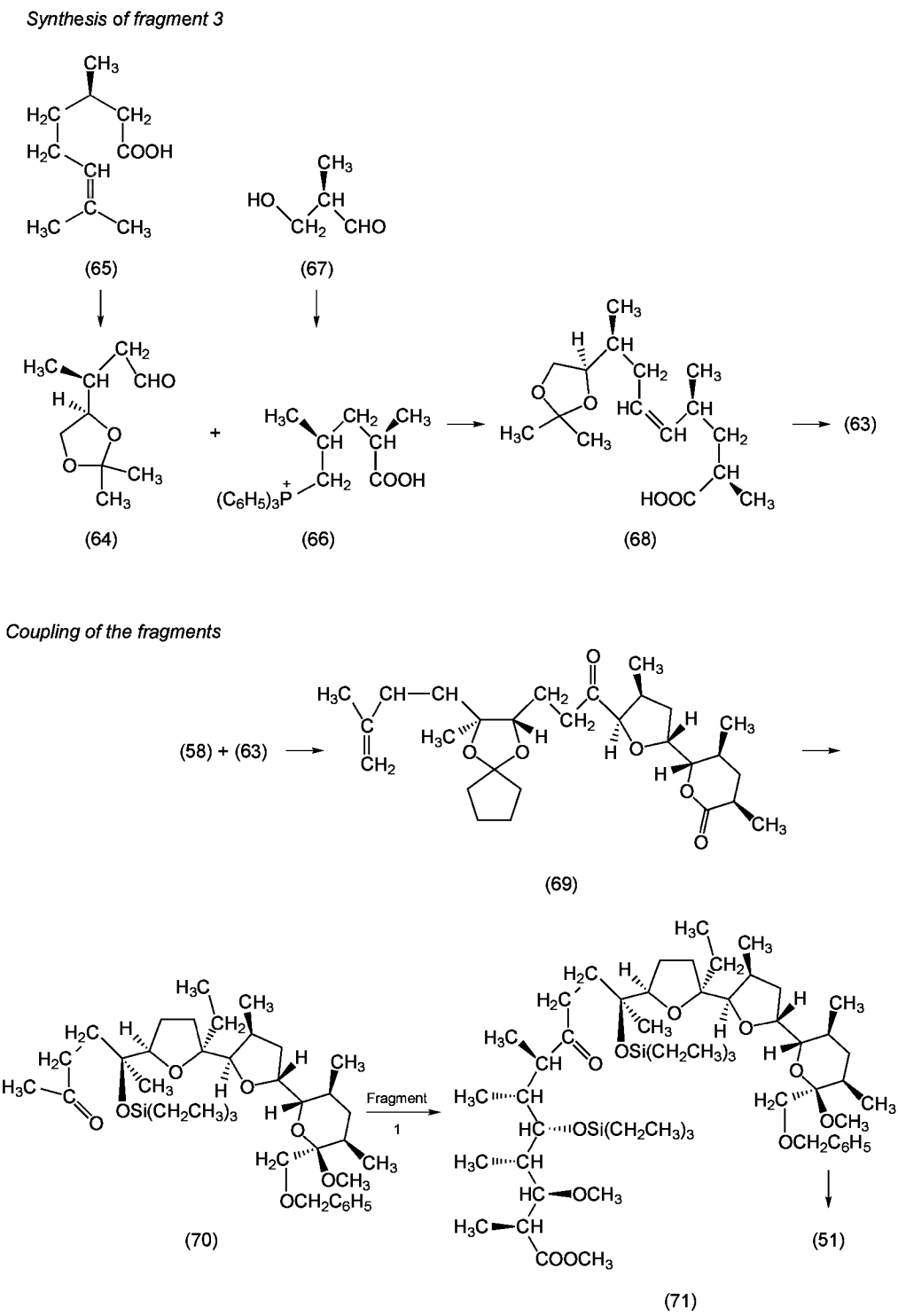
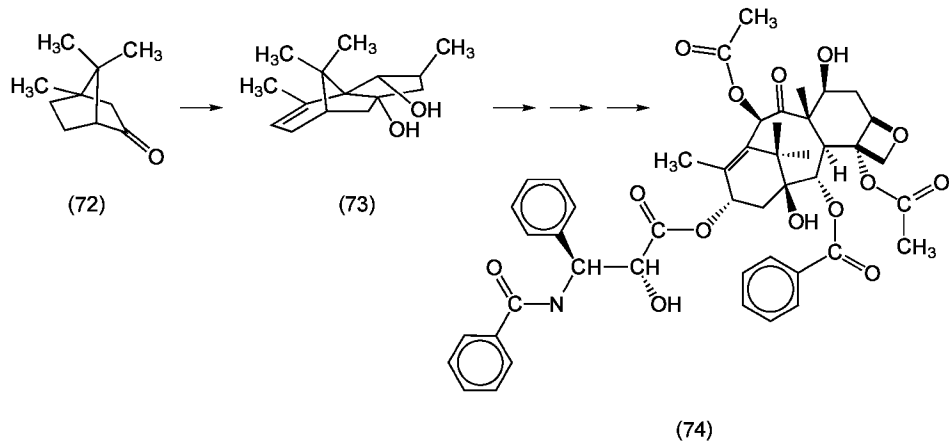
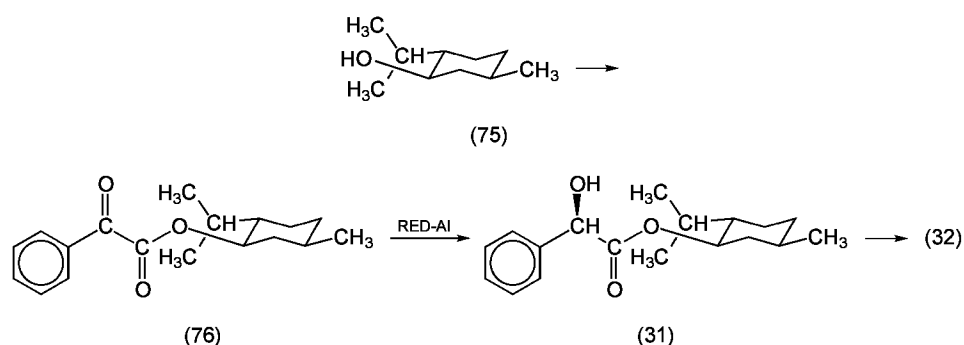


Fig. 4b. *Continued*

Many notable examples of the synthesis of complex natural products from optically pure starting materials have been reported (70). One synthesis of considerable interest is that of taxol [33069-62-4] (74), a potent antitumor agent used clinically. The starting material (73) used in the first total synthesis of taxol is produced in enantiomerically pure form from inexpensive and readily available *l*-camphor [464-48-2] (72) (73).



L-Menthol [2216-51-5] (**75**) and D-menthol [15356-70-4] have been used as chiral auxiliaries in the synthesis of optically active mandelic acids. Reduction of (-)-menthol benzoylformate (**76**) with a sterically bulky reducing agent, ie, sodium bis(2-methylethoxy)aluminum hydride (RED-Al), followed by saponification, yields (R)-mandelic acid (**32**) of 90% ee.



One advantage in using chiral auxiliaries is that they may be recycled. Evan's chiral oxazolidinones, such as (**78**) (Fig. 5), are prepared from readily available amino acids such as L-phenylalanine [63-91-2] (**77**). Reduction of L-phenylalanine with borane followed by reaction with diethylcarbonate produces enantiomerically pure oxazolidinone auxiliary (**78**). The use of this auxiliary is exemplified in the preparation of an intermediate (**82**) in the synthesis of the macrolide antibiotic rutamycin B [1404-59-7] (**83**) (**74**). Reaction of auxiliary (**78**) with butanoyl chloride yields chiral intermediate (**79**). Enolate formation with sodium hexamethyldisilazane produces the sodium-chelated intermediate (**80**). Addition of allyl iodide results in stereoselective alkylation, from the side opposite the protruding benzyl group of the auxiliary, yielding (**81**). Lithium peroxide-catalyzed hydrolysis provides enantiomerically pure 2-ethyl-4-pentenoic acid (**82**) and the chiral auxiliary (**78**).

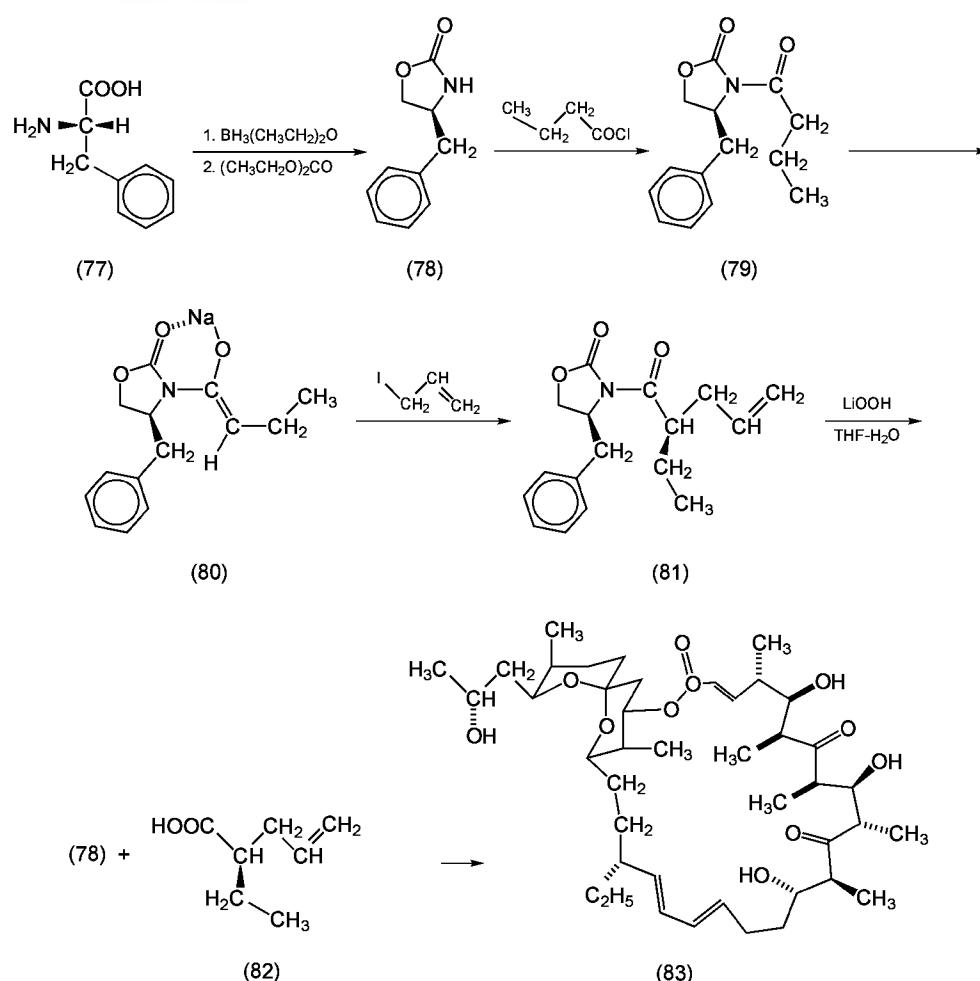


Fig. 5. Synthesis of Evan's chiral oxazolidinone auxiliary and use of this auxiliary for preparation of a chiral intermediate in the synthesis of rutamycin B.

Oppolzer's chiral sultams, such as (-)-bornane-10,2-sultam [94594-90-8] (**85**) (Fig. 6), have been used to control stereoselective enolate formation and direct stereoselective alkylation in the synthesis of numerous nonnatural amino acids (**75**). Both (-) and (+)-bornane-10,2-sultams are prepared from commercially available (+)-(**84**) and (-)-camphorsulfonic acids. The use of a chiral sultam for the synthesis of L-phenylalanine (**90**) is shown in Figure 6. Trimethylaluminum-catalyzed reaction of sultam (**85**) with a glycine methyl ester derivative (**86**) yields the sultam intermediate (**87**). *n*-Butyllithium-assisted enolate formation provides the *Z*-enolate-chelate structure (**88**). Stereoselective reaction with benzylbromide produces the diastereomer (**89**). Imine hydrolysis followed by removal of the auxiliary through hydrolysis (LiOH) produces (**90**) in 93% overall yield with greater than 99.8% ee.

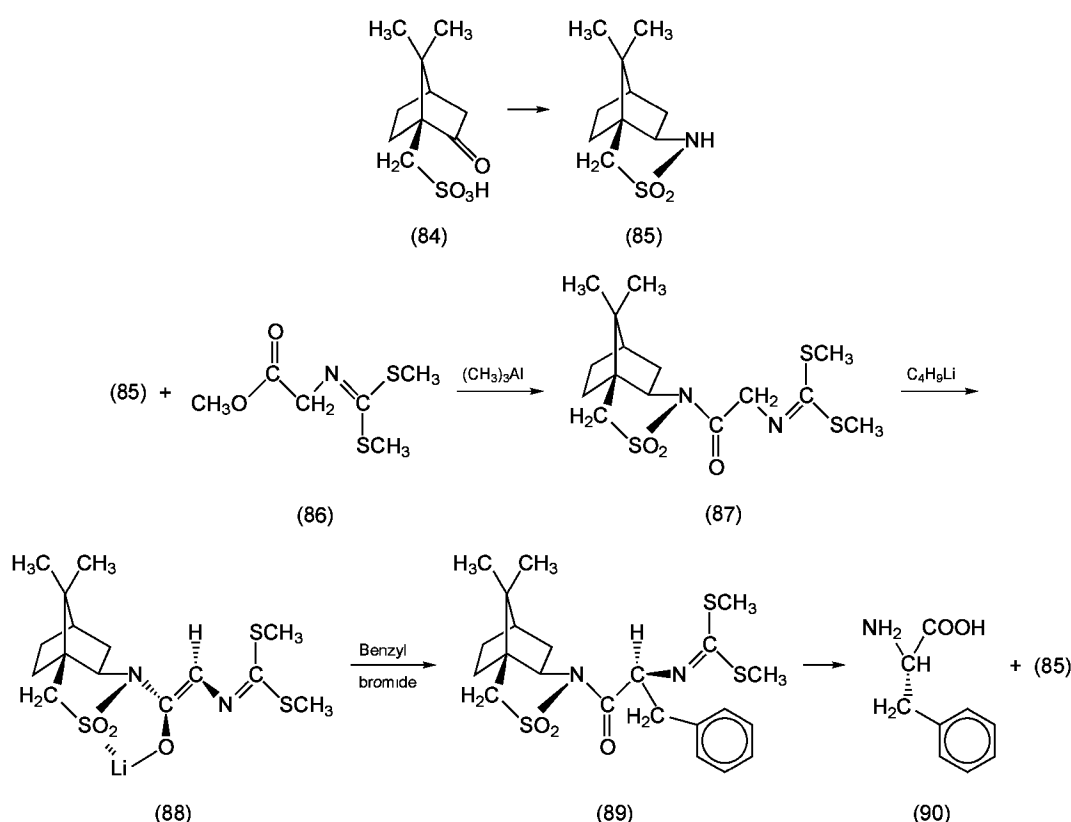


Fig. 6. Preparation of Oppolzer's chiral sultam auxiliary (85) and use of this chiral auxiliary for the preparation of L-phenylalanine.

Asymmetric Induction Methodologies. Asymmetric synthesis is defined as the construction of new chiral centers within a prochiral molecule, with the condition that one optical isomer is formed to a greater extent than the other. The most common type of asymmetric induction involves the conversion of a trigonal carbon atom to a tetrahedral carbon atom by use of a reagent that is biased toward preferential attack from one side or face of the prochiral molecule. Generally, asymmetric induction involves diastereotopic transition states wherein one transition state is favored due to steric and electronic effects which govern the selective formation of one enantiomer. An overview of methods in asymmetric syntheses is available (76). The primary advantage of asymmetric synthesis resides in the stereoselective production, in general, of either enantiomer of a compound; synthesis of both enantiomers is not always possible using chiral synthons or auxiliaries. Asymmetric syntheses avoid the use of inefficient resolutions, and often the reagent or catalyst is recyclable making the synthetic process both material and cost efficient. Asymmetric reduction of ketones, epoxidation of allylic alcohols, hydroboration, hydrogenation, and dihydroxylation reactions, as well as asymmetric cycloaddition reactions, allyl borations, and aldol condensations represent several classes of the numerous enantioselective reactions developed since the 1960s.

Several examples which demonstrate the utility of asymmetric induction in the synthesis of pharmaceutical agents follow. Alpine-borane [42371-63-1] (93) (Fig. 7), a chiral reducing agent useful in the enantioselective reduction of prochiral ketones, is prepared by reaction of 9-borabicyclo-[3.3.1]-nonane [21205-91-4] (91) with either enantiomer of optically pure α -pinene [7785-26-4] (92). This reagent has been employed in the enantioselective reduction of acetylenicketones (94) in better than 85% yield and 92% ee to produce propargyl alcohols (95) which can be converted to 4-substituted butyrolactones (96). Numerous biologically active natural products contain the butyrolactone ring system (77). Similarly, methyl α -aryl- α -ketoacetates (97) have been reduced to methyl (R)- (98) or (S)-mandelates, useful intermediates in the synthesis of, among other compounds, optically active 4-aryl-2-hydroxytetronic acids (99) (78). These acidic compounds are *ac*-reductones which possess a low redox potential and numerous biological activities including reactive radical scavenging, cyclooxygenase inhibition, and antilipidemic properties. The proposed transition state (100) for the enantioselective reduction is a cyclic transition state wherein the boron is coordinated with the ketone carbonyl oxygen and the hydride is delivered from the β -carbon of the pinene (see Fig. 7). The large group of the ketone (R^1) is positioned in the sterically favored equatorial conformation and the small group (R^2) is in an axial position. The cyclic transition state is thought to undergo a sigmatropic-type rearrangement in which a hydride is delivered to the carbonyl carbon, α -pinene (92) is generated, and a boronate ester (101) is formed.

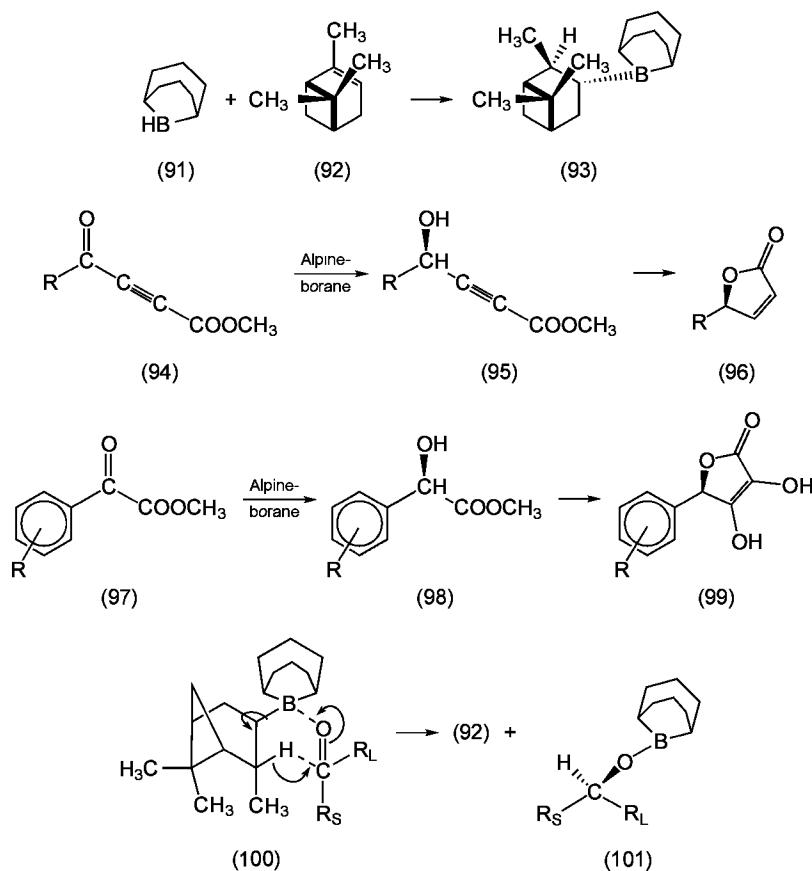


Fig. 7. Preparation of Alpine-borane (93) and use in the synthesis of homochiral butyrolactones and arylhydroxytetronic acids. R_s and R_L denote small and large groups, respectively.

Asymmetric epoxidation is of high utility; two stereocenters may be constructed simultaneously. In general, allylic alcohols are oxidized in high yields and with high enantiomeric purity using titanium tetrakisopropoxide and enantiomerically pure diisopropyl tartrate as catalysts, dichloromethane as solvent, and *tert*-butyl hydroperoxide as the oxidant (79) (Fig. 8). The starting epoxide (103) used in the enantiomeric synthesis of the immunosuppressant FK-506 (105) is prepared by Sharpless epoxidation of 1,4-pentadiene-3-ol [922-65-6] (102) (80). The proposed intermediate (104), which gives rise to the enantiofacial selectivity, involves a complex of diisopropyl tartrate, *tert*-butyl hydroperoxide and the allylic alcohol all bonded to the titanium(VI) species (81).

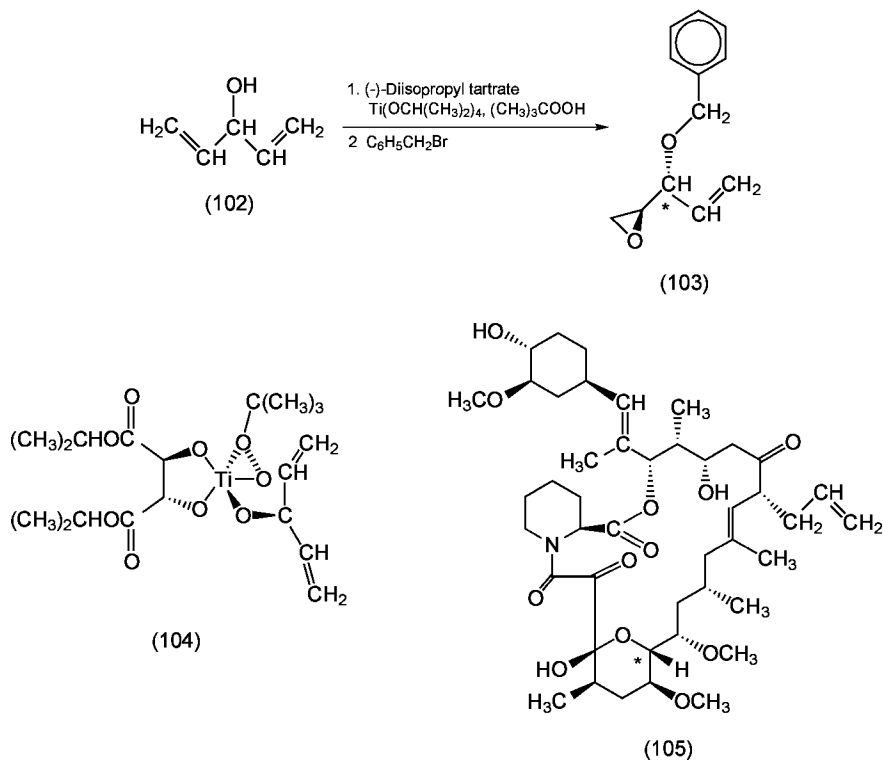


Fig. 8. Use of Sharpless asymmetric epoxidation for the preparation of an intermediate in the synthesis of FK-506 (105), where * represents the chiral carbon of (103).

Catalytic asymmetric hydrogenation was one of the first enantioselective synthetic methods used industrially (82). 2,2'-Bis(diarylphosphino)-1,1'-binaphthyl (BINAP) is a chiral ligand which possesses a C₂ plane of symmetry (Fig. 9). Steric interactions prevent interconversion of the (R)- and (S)-BINAP. Coordination of BINAP with a transition metal such as ruthenium or rhodium produces a chiral hydrogenation catalyst capable of inducing a high degree of enantiofacial selectivity (83). Naproxen (41) is produced in 97% ee by Ru(OCOCH₃)₂[(S)-BINAP]-(106)-catalyzed reduction of precursor olefin (107). The asymmetric synthesis of analgesic tetrahydroisoquinolines makes use

of this methodology (84). L-3,4-Dihydroxyphenylalanine [59-92-7] (L-dopa) (**109**), useful in the treatment of Parkinson's disease, is prepared by asymmetric catalytic hydrogenation of intermediate olefinic amide (**108**).

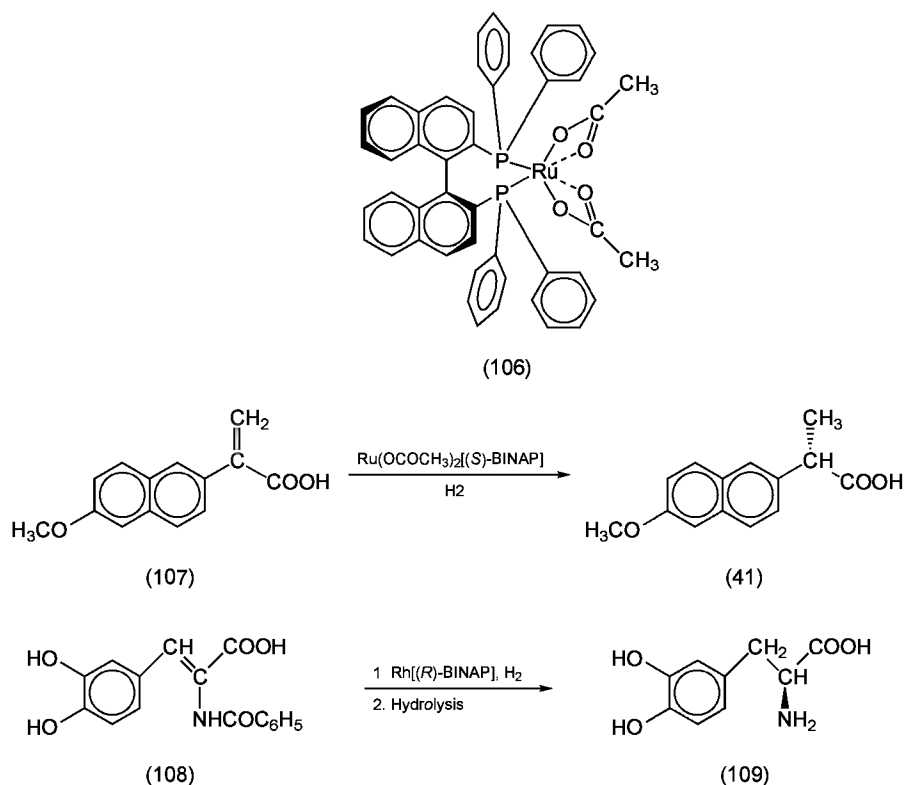
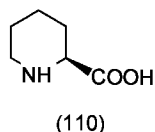


Fig. 9. Catalytic asymmetric hydrogenation.

Analysis of Synthetic Homochiral Drugs

Determination of Absolute Configuration. X-irradiation of a crystal produces a diffraction pattern from which the relative spatial orientation of the atoms that make up the molecule may be determined. If the crystal is made of homochiral molecules, the absolute configuration of the compound may be deduced. In 1951 the absolute configuration of NaRb-(+)-tartrate was determined to be L-(+) (or R)-(+)) (85). Subsequently, the absolute stereochemistry of (+)-glyceraldehyde (**5**) was deduced to be of the D-configuration by chemical correlation with L-(+)-tartaric acid. Fischer's original arbitrary assignment of D-(+)-glyceraldehyde is structurally correct; consequently the structures of the numerous compounds deduced from D-glyceraldehyde are also correct.

Chemical conversion of compounds to intermediates of known absolute configuration is a method routinely used to determine absolute configuration (86). This is necessary because x-ray analysis is not always possible; suitable crystals are required and determination of the absolute configuration of many crystalline molecules cannot be done because of poor resolution. Such poor resolution is usually a function of either molecular instability or the complex nature of the molecule. For example, the relative configuration of the macrolide immunosuppressant FK-506 (**105**) (Fig. 8), which contains 14 stereocenters, was determined by x-ray crystallographic studies. However, the absolute configuration could only be elucidated by chemical degradation and isolation of L-pipecolic acid (**110**) (80).

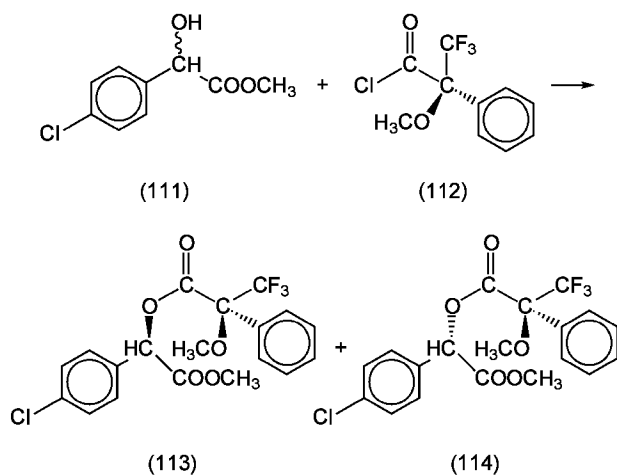


ORD and CD also provide a basis by which the absolute configuration of a compound may be correlated with that of a known compound of similar structure by observing changes in degree of rotation with wavelength (87).

Determination of Enantiomeric Purity. In order to analyze the biological properties of a single enantiomer, the optical purity of the compound should be enantiomerically pure, ie, 100% ee. Contrasting reports on the differences in pharmacological activity of single enantiomers, as well as the misinterpretation of data, are often a result of unknowingly testing enantiomerically impure material (88). The oldest and perhaps easiest method for determining optical purity is by measuring optical rotation and comparing the value with that reported for the enantiopure compound. Although simple, there are several drawbacks to this method. The assumption must be made that the reported literature value is without error, and truly represents the optically pure compound. Numerous examples exist in which unambiguous methods, ie, chiral gc, hplc, nmr, for determination of optical purity reveal that the previously reported values for optically pure compounds were in error. Variables such as temperature, solvent, concentration, purity of the compound, type of cell, and even differences between polarimeters employed in the measurement influence the observed degree of rotation. Therefore, polarimetry measurements for determination of optical purity deviate by at least ±4% (89).

¹H-nmr is commonly used to determine enantiomeric purity and is reliable to above 98% ee. Chiral shift reagents are employed to separate the resonance signals of enantiomers (90). Chiral shift reagents exemplified by tris(dipivaloylmethanato)europium, Eu(dpm)₃, act as weak Lewis acids and their association with organic compounds results in the spreading or separation of their proton resonance signals. Furthermore, the association of the chiral shift reagent creates a diastereotopic environment resulting in the resolution of the proton resonance signals for the individual enantiomers.

Diastereomeric derivatization of a chiral alcohol (**111**) with an enantiopure compound such as Mosher's reagent [20445-33-4] (α-trifluoromethyl-α-methoxy-α-phenylacetylchloride) (**112**) (91) results in two distinct compounds (**113**) and (**114**) with nonequivalent chemical shifts in the ¹H-nmr spectrum (92).



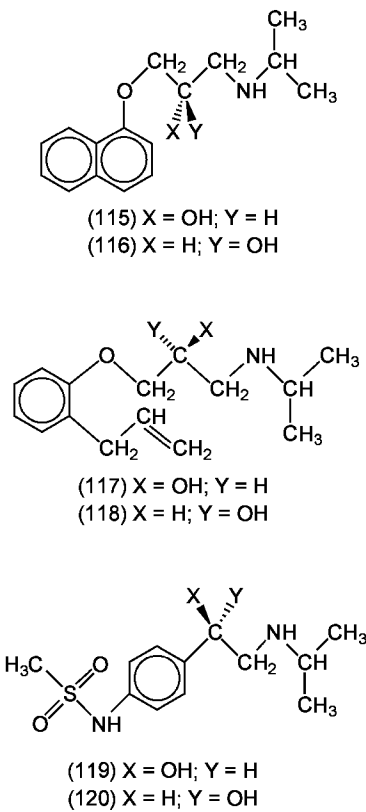
Integration of the peaks for the two diastereomers accurately quantifies the relative amounts of each enantiomer within the mixture. Such diastereomeric derivatives may also be analyzed by more accurate methods such as gc or hplc. One drawback to diastereomeric derivatization is that it requires at least 15 mg of material, which is likely to be material painstakingly synthesized, isolated, and purified. The use of analytical chiral chromatographic methods allows for the direct quantification of enantiomeric purity, is highly accurate to above 99.8% ee, and requires less than one milligram of material.

Chiral Pharmaceuticals

ENANTIOMERIC PAIRS

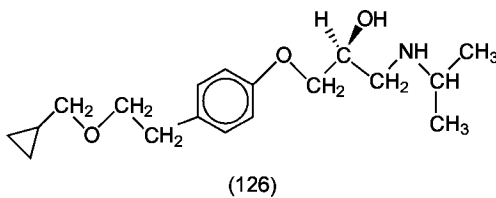
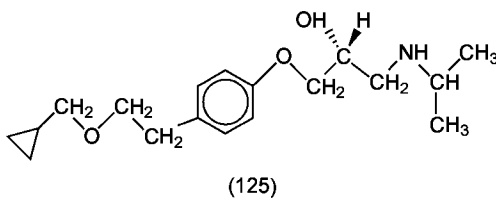
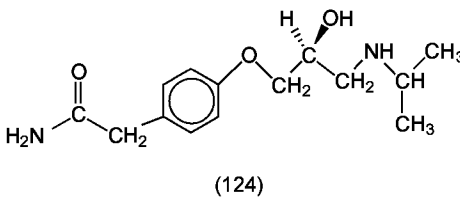
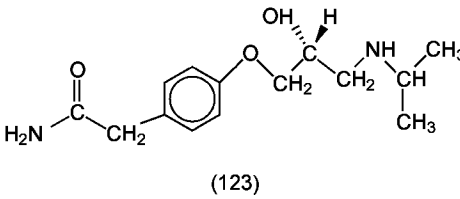
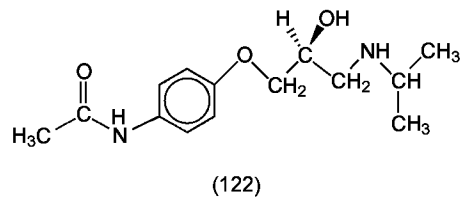
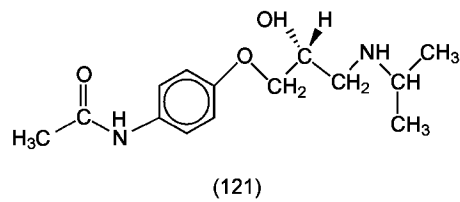
Enantioselective differences in absorption, metabolism, clearance, drug– macromolecule binding affinity, and other factors, which culminate in the observed enantioselective efficacy of chiral drugs, are considered herein. More inclusive lists of optically active drugs and their enantioselective differences are available (93).

Antihypertensive Agents. Hypertension (high blood pressure) is a significant risk factor for cardiovascular diseases such as angina, heart attacks, and strokes. β -Adrenoceptor (adrenergic nervous system receptors of the β -type) antagonists (β -blockers), calcium channel blockers, angiotensin-converting enzyme (ACE) inhibitors, and potassium channel activators (KCAs) are among the numerous classes of drugs developed to control hypertension (see CARDIOVASCULAR AGENTS; ENZYME INHIBITORS; NEUROREGULATORS). β -Adrenoceptor antagonists exemplified by the phenoxypropanolamine derivatives propranolol (115) and (116) and alprenolol (117) and (118) or the phenethanolamine drugs such as sotalol (119) and (120) require for activity both the ethanolamine portion and an aromatic ring. Furthermore, the correct spatial arrangement of the phenyl, ethylamine, and hydroxyl moieties is critical for β -blockade (94). For example, in guinea pig atria, the (R)-enantiomer of alprenolol [23846-72-2] (118) is over 100-fold more active than is the (S)-enantiomer [23846-71-1] (117). Similarly, the (S)-enantiomer of sotalol [30236-32-9] (120) is 50 times as potent as (R)-sotalol [30236-31-8] (119) (95).

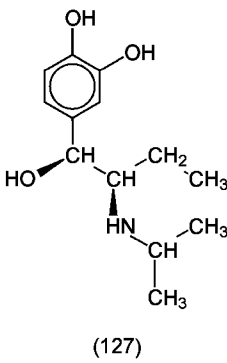


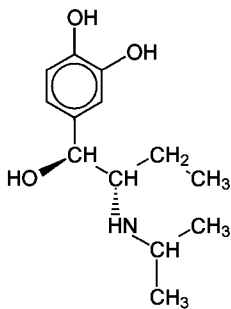
Note that the relative spatial arrangement of the phenyl, amine, and hydroxyl functionalities are identical for (R)-alprenolol and (S)-sotalol. In addition to β -blocking activities, some of these compounds also possess potent local anaesthetic activity (see ANESTHETICS). The membrane stabilizing activity, however, is not stereoselective and correlates directly with the partition coefficient (hydrophobicity) of the compound.

Two different types of β -adrenoceptors have been characterized and categorized as β_1 - and β_2 -subtypes. The β_1 -receptors are associated primarily with the cardiac muscle, whereas the β_2 -subtype is located peripherally. Selective β_1 -blockers include practolol (121) and (122), atenolol (123) and (124), and betaxolol (125) and (126).

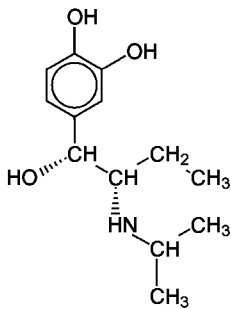


It has been demonstrated that the β_1 -selectivity is due to the para-substituents of these drugs (96). In contrast, (-)-erythro-isoetharine (**127**), a bronchodilator, is 80 times more selective for β_2 -adrenergic receptors than for β_1 -receptors. Isoetharine (97) contains an α -alkyl substituent, thus producing four isomeric compounds. The ()-erythro isomer (**127**) is 100-fold more active than the ()-threo isomer (**128**) and has more than 500 times the activity of either of the ()-isomers (**129**) and (**130**) in blocking electrically stimulated spasms in guinea pig trachea. In general, introduction of α -alkyl substituents on both β -blockers and agonists provides diastereomers with increased β_2 -selectivity, but often with compromised potency.

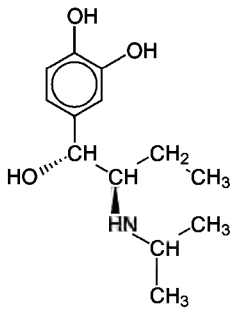




(128)

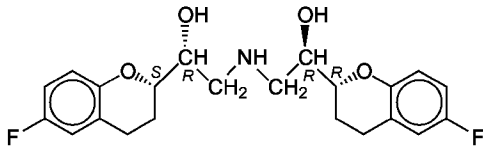


(129)

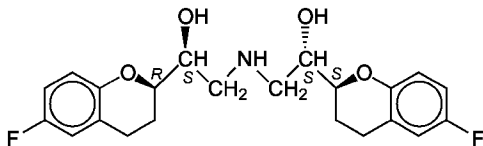


(130)

Racemic nevirapine (**131**) and (**132**), a nonclassical β -blocker, contains four stereocenters with the (*S*,*R*,*R*,*R*)-(+)-diastereomer (**131**) being a long-acting, potent, and highly selective β_1 -blocker (50-fold more selective for β_1 versus β_2). The (–)-enantiomer (**132**) does not possess significant β_1 -activity. When nevirapine is administered in racemic form sharp decreases in both diastolic and systolic blood pressure, attributable to the (*R*,*S*,*S*,*S*)-(–)-diastereomer (**132**), are observed. The hemodynamic effects cannot be explained by β_1 -adrenergic antagonism of the (+)-isomer and this effect is not shared by other β -blockers. The mechanism of the observed synergy between the two enantiomers is not known (98).



(131)



(132)

α -Adrenergic adrenoceptors (99–101) exist in two isoforms designated α_1 and α_2 . Both subtypes are observed in equal concentrations post-junctionally on vascular smooth muscle, while the α_2 -subtype occurs more frequently at the presynaptic junction (102). α -Adrenoceptor agonists, which induce vasoconstriction, include the phenethylamines noradrenaline (**133**) and (**134**) (Fig. 10) and α -methylnoradrenaline (**9,10**). The diastereomeric (1*R*,2*S*)-(-)- α -methylnoradrenaline erythro isomer (**136**) (103) stereoselectively binds 550-fold more tightly to the α_2 -receptor subtype than to the α_1 -receptor subtype in guinea pig ileum. The other three isomers are relatively inactive. The relative biological activities of two of the noradrenaline diastereomers and dopamine, a related compound, are (*R*)-(-)-noradrenaline (**133**) > (*S*)-(+)-noradrenaline (**134**) > dopamine (**135**). The receptor subtype selectivity, as well as the stereoselectivity observed with these agonists, has been explained by the three-point interaction hypothesis (104,105). This hypothesis suggests that only one enantiomer is capable of existing in a conformation (Fig. 10) such that favorable interactions exist between the receptor and the cationic amine, substituted phenyl, and β -hydroxyl groups of the agonist. The hypothesis further states that symmetrical phenethylamines such as dopamine which are devoid of a β -hydroxyl moiety, should be equal in activity to the less active enantiomer of the β -hydroxyphenethylamines. This hypothesis successfully predicts the stereoselectivity of both agonists and antagonists of the α -adrenoceptors and the β -adrenoceptors (106).

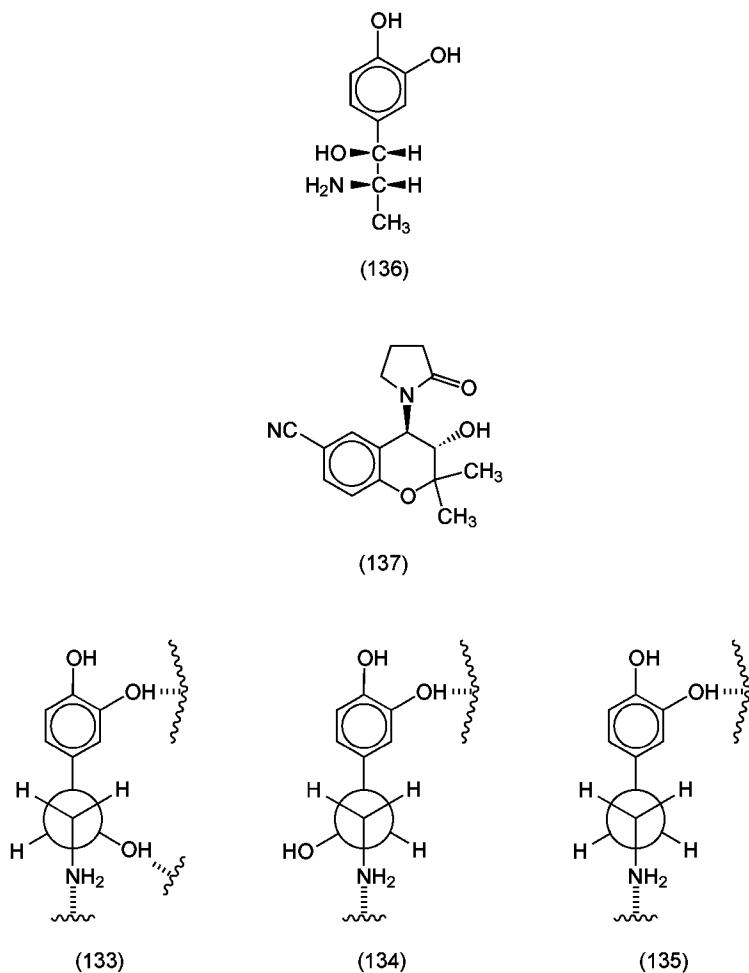
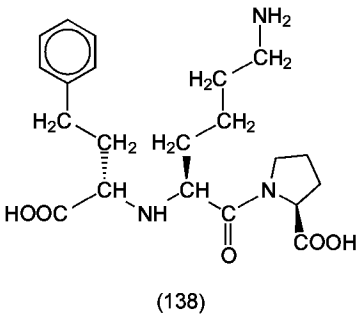


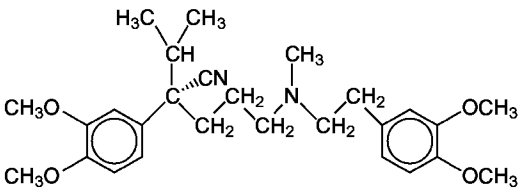
Fig. 10. The postulated interaction of α-adrenoceptor agonists with the receptor. The Easson-Stedman hypothesis suggests that (R)-noradrenaline is most potent owing to its three points of attachment (~) to the adrenoceptor, whereas dopamine and (S)-noradrenaline are equal in activity, but less active than (R)-noradrenaline because they each possess two binding domains (100).

Cromakalim (137) is a potassium channel activator commonly used as an antihypertensive agent (107). The rationale for the design of cromakalim is based on β-blockers such as propranolol (115) and atenolol (123). Conformational restriction of the propanolamine side chain as observed in the cromakalim chroman nucleus provides compounds with desired antihypertensive activity free of the side effects commonly associated with β-blockers. Enantiomerically pure cromakalim is produced by resolution of the diastereomeric (S)-α-methylbenzylcarbamate derivatives. X-ray crystallographic analysis of this diastereomer provides the absolute stereochemistry of cromakalim. Biological activity resides primarily in the (–)-(3S,4R)-enantiomer [94535-50-9] (137) (108). In spontaneously hypertensive rats, the (–)-(3S,4R)-enantiomer, at dosages of 0.3 mg kg, lowers the systolic pressure 47%, whereas the (+)-(3R,4S)-enantiomer only decreases the systolic pressure by 14% at a dose of 3.0 mg kg.

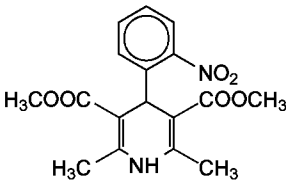
Angiotensin converting enzyme (ACE) inhibitors alleviate hypertension by blocking the endogenous synthesis of angiotensinII via the renin pathway. ACE is a zinc-containing carboxypeptase which cleaves the His–Leu dipeptide from the C-terminal end of angiotensin I. Captopril (47), 1-[(2S)-3-mercapto-2-methylpropanoyl]-(2S)-proline, is a prototype ACE competitive inhibitor with a Ki of 1.7 nM. The diastereomer 1-[(2R)-3-mercapto-2-methylpropanoyl]-(2S)-proline is 100 times less active than captopril (109). Lisinopril [83915-83-7] (138), a dipeptide containing an (S)-lysine, is more than five times as potent as the epimeric mixture consisting of (R) + (S)-lysine moieties in inhibiting hog plasma ACE (110).



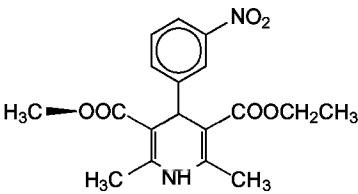
Calcium channel antagonists such as verapamil (139), diltiazem, prenylamine, and the 1,4-dihydropyridines nifedipine [21829-25-4] (140), nitrendipine (141) and (142), and nimodipine (143) and (144) are effective in the treatment of angina (111). The 1,4-dihydropyridine, nifedipine (140), is not chiral owing to the symmetry of the dihydropyridine ring system. However, replacement of one of the methyl esters with a different substituent introduces asymmetry into the molecule and enantioselective calcium antagonism is observed. (–)-Nitrendipine [80873-62-7] (142), consisting of methyl and ethyl esters, is 10 times more potent in rabbit aorta than is the (+)-isomer (141) (112). When the esters are methyl and isopropyl, the (+)-isomer (145) is 100 times more effective than the (–)-isomer (146). However, the (–)-enantiomer (146) is 10-fold more potent on the guinea pig ileum (112). (S)-Verapamil [38321-02-7] (139) produces vasodilation 2.5 times better than its enantiomer. (S)-Verapamil, however, is metabolized in the liver faster than its mirror image isomer after oral dosing. This generally results in 3 to 10 times more (R)-verapamil [36622-29-4] in the systemic circulation.



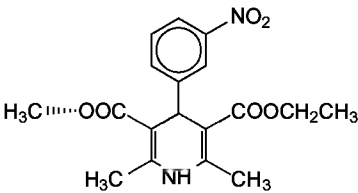
(139)



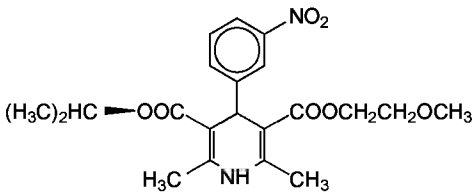
(140)



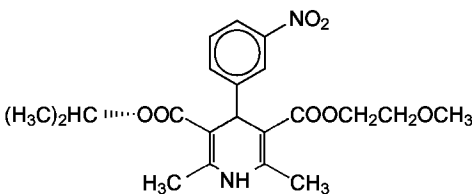
(141)



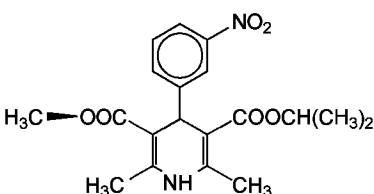
(142)



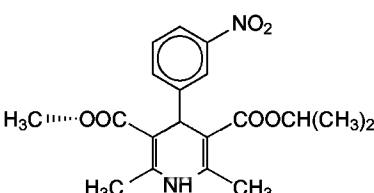
(143)



(144)

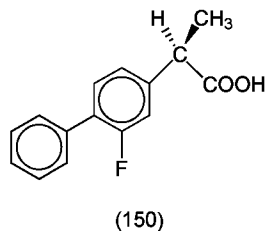
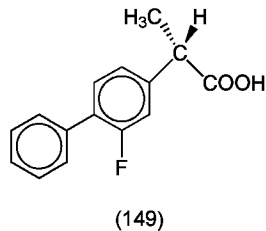
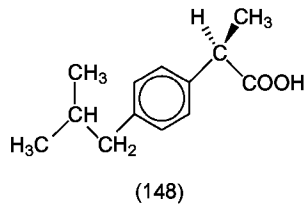
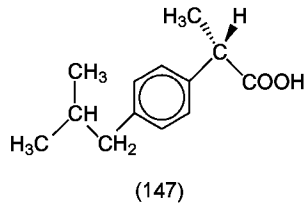


(145)



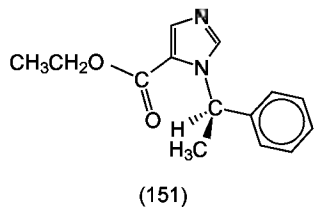
(146)

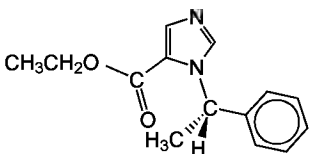
Nonsteroidal Antiinflammatory Drugs. Nonsteroidal antiinflammatory drugs (NSAIDs) include, among the numerous agents of this class, aspirin (acetylsalicylic acid), the arylacetic acids indomethacin and sulindac, and the arylpropionic acids, (*S*)-**(147)** and (*R*)-**(148)** ibuprofen, (*S*)-**(149)** and (*R*)-**(150)**, flurbiprofen naproxen **(41)**, and fenoprofen (see ANALGESICS, ANTIPYRETICS, AND ANTIINFLAMMATORY AGENTS; SALICYLIC ACID AND RELATED COMPOUNDS).



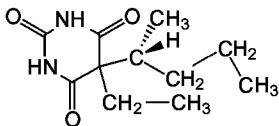
Although the arylpropionic acids contain a stereogenic center they are generally marketed as racemic mixtures. The only exception is naproxen **(41)**, which is marketed as its (*S*)-enantiomer. NSAIDs produce their antiinflammatory effects by inhibiting cyclooxygenase (COX), the enzyme which catalyzes the first transformation in the biosynthetic conversion of arachidonic acid to the 20 carbon prostaglandins. The (*S*)-arylpropionic acids are the active enantiomers. (*S*)-(+)-Ibuprofen **(147)** is 160 times more potent than its enantiomorph *in vitro*. However, although (*R*)-(-)-ibuprofen **(148)** is inactive *in vitro*, there is no difference between the antiinflammatory activity of the two enantiomers *in vivo*. The activity *in vivo* of the (*R*)-enantiomorph is due to the enzyme mandelate racemase, which selectively converts inactive (*R*)-ibuprofen into the active (*S*)-enantiomer. Two COX isoforms, COX-1 and COX-2, have been identified, and the crystal structure of COX-1 containing (*S*)-flurbiprofen **(149)** has been elucidated. The carboxylate anion of the drug forms an ionic bond with Arg-243 of the cyclooxygenase. The distal flurbiprofen phenyl ring appears to be associated with the phenyl ring of Tyr-245; the methyl group projects into a hydrophobic pocket. The lower activity of the (*R*)-enantiomer **(150)** is likely to be due to unfavorable interactions between its methyl group and the enzyme.

CNS Depressant Drugs. Central nervous system (CNS) depressant drugs (113) including antianxiety agents (benzodiazepines) (114), sedative–hypnotics, general anesthetics, and certain spasticity agents all demonstrate high degrees of enantioselective activity (see HYPNOTICS, SEDATIVES, ANTICONVULSANTS, AND ANXIOLYTICS; PSYCHOPHARMACOLOGICAL AGENTS). (*R*)-(+)-Etomidate [33125-97-2] **(151)** is a short-acting and potent hypnotic, whereas the (*S*)-(-)-isomer **(152)** is devoid of hypnotic activity. The exact mechanism of action of etomidate is speculative; however, the observed enantioselectivity provides evidence that a receptor is involved. Brain levels of each enantiomer are equal. Barbiturates are commonly prescribed for their sedative–hypnotic activities. In general, the (*S*)-(-)-enantiomers possess CNS depressant activities, whereas the (*R*)-(+)-isomers often produce an excitatory effect (115). In humans, (*R*)-(+)-pentobarbital [21045-50-1] **(153)** is found bound to human plasma proteins to a lesser extent than the (*S*)-(-)-isomer **(154)** (36.6% free vs 26.5% free) and is subsequently cleared 14% faster (116). This increased rate of clearance is not sufficient to account for the two- to threefold greater duration of action of (*S*)-(-)-pentobarbital [5767-32-8] **(154)**, and suggests that the difference in activity between the enantiomers is due to the pharmacodynamics of the more potent (*S*)-isomer. (*S*)-(+)-Hexobarbital [7245-04-7] **(155)**, the eutomer, is eliminated about 2.5 times more slowly than the inactive (*R*)-(-)-isomer **(156)**, a result of differences in hepatic metabolism (115,117). Diazepam [439-14-5] **(157)**, an achiral benzodiazepine, undergoes stereoselective metabolism to (*S*)-(+)-oxazepam [52432-56-1] **(158)** in the liver (118). (*S*)-(+)-Oxazepam produces antianxiety effects to a greater degree than the mirror image isomer **(159)**.

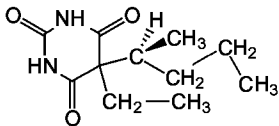




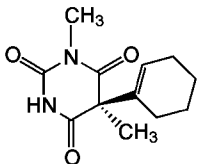
(152)



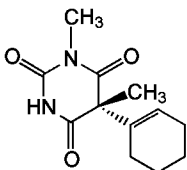
(153)



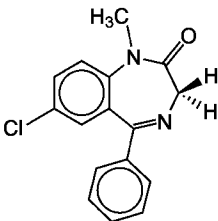
(154)



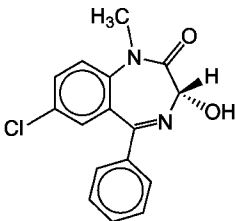
(155)



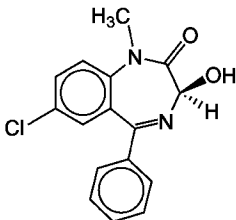
(156)



(157)



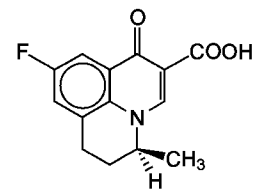
(158)



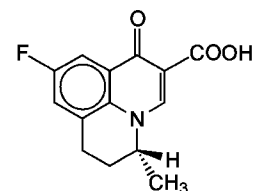
(159)

Antibiotic and Antimicrobial Drugs. The antimicrobial agents (119) flumequine **(160)** and **(161)**, and methylflumequine (S-25930) **(162)** and **(163)** effectively eliminate a number of microbial pathogens via inhibition of the topoisomerase II enzyme of c-DNA containing bacteria (120) (see

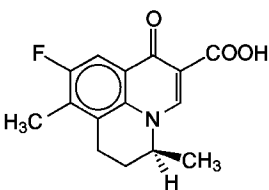
ANTIBACTERIAL AGENTS, SYNTHETIC). The (*S*)-enantiomers (**160**) and (**162**) of both drugs are much more potent than the (*R*)-enantiomers (**161**) and (**163**). The potent analogue (*S*)-(-)-ofloxacin [82419-36-1] (**164**) is 8–125 times more potent than its enantiomer (**165**) (121) although it is sold only as the racemate. In humans the disposition of (*R*)- and (*S*)-enantiomers of ofloxacin is stereoselective due to differences in renal clearance rates (122). This difference, however, does not fully explain the large enantioselective difference in antibacterial potency. β-Lactam antibiotics (see ANTIBIOTICS, β-LACTAMS), such as the penicillins and cephalosporins, require the (3*S*,5*R*,6*R*)-configuration of the β-lactam functionality combined with a D-amine in either the 6-position (penicillins) or 7-position (cephalosporins) to produce optimal activity (119). The β-lactam antibiotic 7-*L*-cephalexin (**166**) is stereoselectively absorbed across the intestinal mucosa via the dipeptide transport system. However, only the 7-*D*-cephalexin [15686-71-2] (**167**) epimer is observed in serum and urine after oral administration owing to rapid and highly stereoselective enzymatic hydrolysis of the *L*-epimer (**166**) (123).



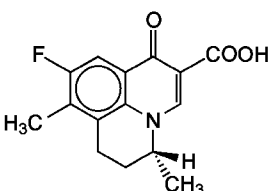
(160)



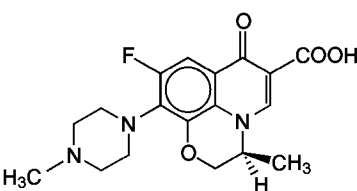
(161)



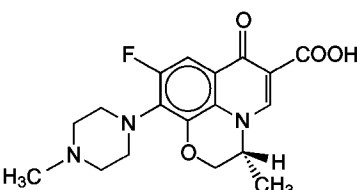
(162)



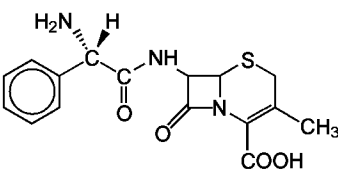
(163)



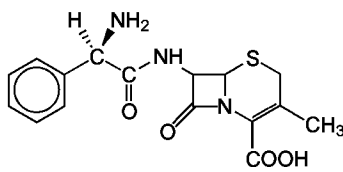
(164)



(165)

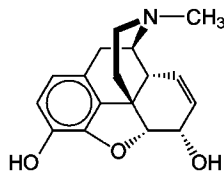


(166)

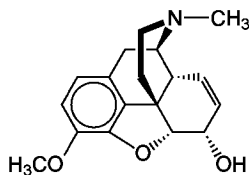


(167)

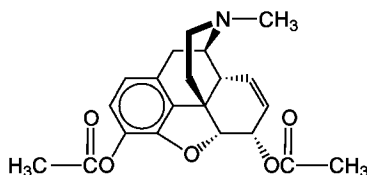
Opioid Analgesic Drugs. (5*R*,6*S*,9*R*,13*S*,14*R*)-(-)-Morphine [57-27-2] (**168**) and its closely related relatives (-)-codeine [76-57-3] (**169**) and (-)-heroin [561-27-3] (**170**) are potent analgesics, while their (+)-isomers possess no analgesic effects (124). α -Dextropropoxyphene (DARVON) (**171**) is a marketed analgesic whereas its enantiomer, α -levopropoxyphene (NOVRAD) (**172**) is sold as an antitussive devoid of analgesic activity (125) (see EXPECTORANTS, ANTITUSSIVES, AND RELATED AGENTS). These analgesics produce their biological effects via stimulation of the opioid receptor subclasses μ -, δ -, κ -, and σ . 5-(*m*-Hydroxyphenyl)-2-methylmorphin (**173**) and (**174**) maintains the basic pharmacophore of morphine: a *m*-hydroxyphenyl functionality bonded to a quaternary carbon containing a tertiary aminoethyl moiety. The (1*R*,5*S*)-(-)-enantiomer (**174**) binds weakly to μ -receptors and is similar to morphine in potency with respect to pain relief. This isomer, however, does not support characteristic opioid dependence in monkeys or rats. In contrast the (1*S*,5*R*)-(+)-isomer (**173**), binds strongly to the μ -receptor and is similar to morphine in its biological effects, including dependence (126). (-)-Methadone (**175**) enantioselectively binds to opiate receptors in rat brain, produces respiratory depression in humans, and blocks serotonin uptake; it is a 30-fold more potent analgesic in rats than is (+)-methadone (**176**) (127). *N*-Allylnormetazocine (NANM) enantiomers (**177**) and (**178**) bind selectively to different receptors in mouse brain. The (-)-enantiomer (**177**), which selectively binds to the μ -opiate receptor with a K_D of 2.1 nM, is a weak agonist in antinociceptive assays and a formidable narcotic antagonist. The (+)-NANM isomer (**178**) binds selectively to the PCP opioid receptor with a K_D of 12 nM and does not compete with its (-)-isomer for the μ -receptor (128).



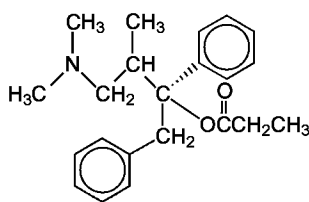
(168)



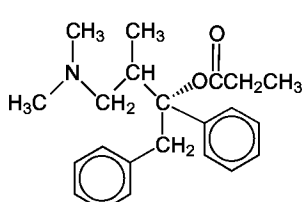
(169)



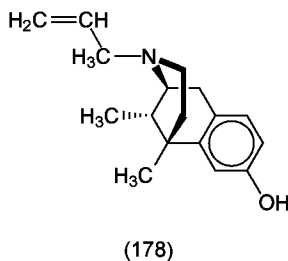
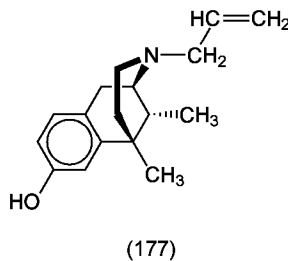
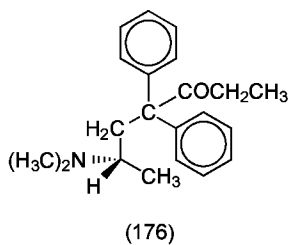
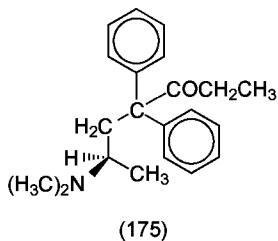
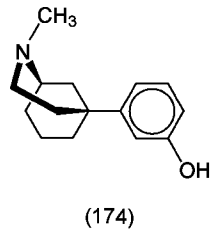
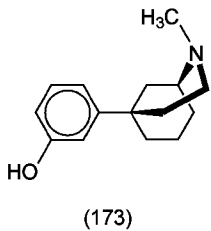
(170)



(171)

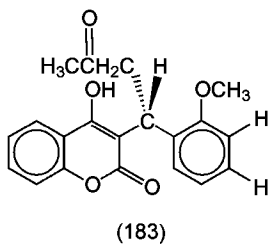
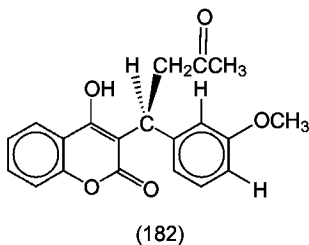
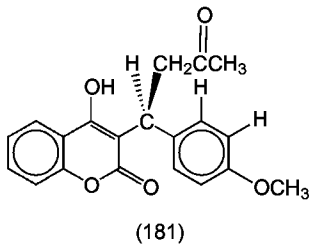
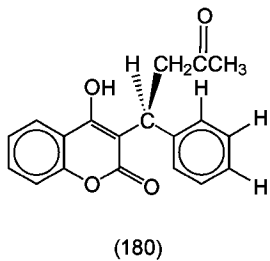
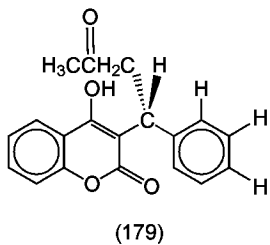


(172)

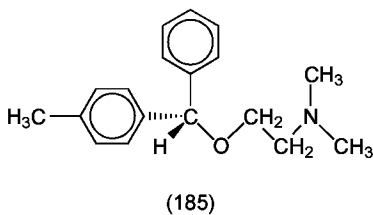
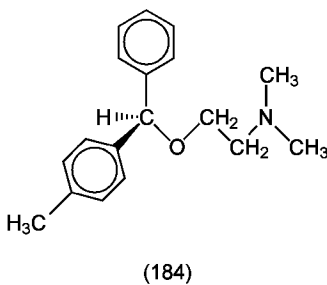


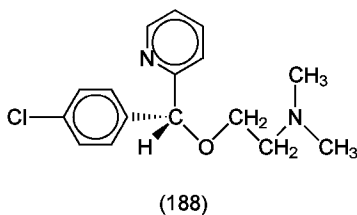
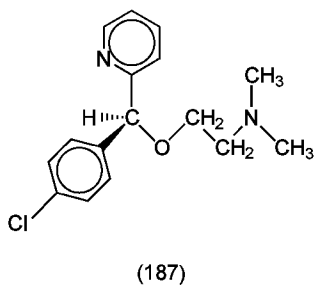
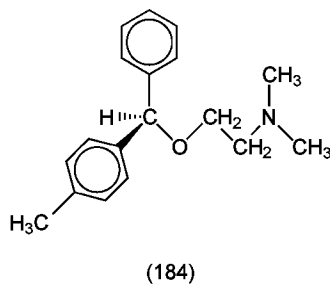
Opioid peptides include the pentapeptides Met-enkephalin (Tyr¹–Gly²–Gly³–Phe⁴–Met⁵) and Leu-enkephalin (Tyr¹–Gly²–Gly³–Phe⁴–Leu⁵) β-endorphin, a 31-amino acid peptide which is a long-lasting analgesic, and the dynorphins which are highly selective and potent kappa-receptor ligands (129) (see OPIOIDS, ENDOGENOUS). Numerous analogues of these opioid receptor ligands have been synthesized and considerable knowledge of the stereochemical binding requirements of the receptors has been acquired. A positively charged N-terminal tyrosine residue is critical for enkephalin activity. Substituents at the C-terminal end such as esters or amides do not significantly alter enkephalin opioid activity, although activity decreases with hydrophilic substituents. Replacement of L-Tyr¹ or L-Phe⁴ with D-Tyr¹ or D-Phe⁴ results in loss of activity. However, substitution of D-aliphatic amino acids for Gly² provides compounds with improved opioid activity owing, in part, to decreased enzymatic hydrolysis of the terminal Tyr¹ residue (130). Dermorphin, a heptapeptide (H-Tyr¹–D-Ala²–Phe³–Gly⁴–Tyr⁵–Pro⁶–Ser⁷–NH₂) isolated from the skin of a South American frog, is considerably more potent than morphine (2000-fold in the hot-plate tail flick test), β-endorphin, or endogenous enkephalins in blocking electrically stimulated contractions of guinea pig ileum and mouse vas deferens *in vitro*. Dermorphin contains D-Ala at position 2; the all L-heptapeptide is 100 times less active (131).

Anticoagulant Drugs. Warfarin (179) and (180), a potent anticoagulant, was first isolated from spoiled clover hay and identified as the agent responsible for the hemorrhagic symptoms associated with the death of livestock in the 1930s. This functionalized coumarin derivative exerts its effect via competitive inhibition of vitamin K-dependent carboxylation of blood clotting factors. Warfarin is generally administered as the racemate, even though (*S*)-warfarin [5543-57-7] (180) is fivefold more active than (*R*)-warfarin [5543-58-8] (179) in both rats and humans (132). The (*S*)-enantiomer is eliminated at a higher rate than its antipode in humans, but in rats, (*R*)-warfarin is more rapidly eliminated. (*S*)-Warfarin and its 3'- and 4'-substituted derivatives, ie, methoxy derivatives (181) and (182), demonstrate stereoselective serum albumin protein binding. However, the (*R*)-2'-substituted warfarin derivative (183) undergoes a greater degree of protein binding than its enantiomer (133). Co-administration with the antiinflammatory drug phenylbutazone increases warfarin's anticoagulant properties (134).

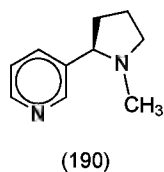
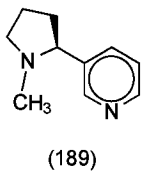


Neurotransmitters. Histamine receptors are found in at least two subtypes designated H₁ and H₂ (see HISTAMINE AND HISTAMINE ANTAGONISTS). H₁-receptor antagonists produce vasoconstriction, while H₂-receptor antagonists inhibit gastric secretion. Neobenodine (**184**) and (**185**), the chiral *p*-methylphenyl analogue of benadryl [58-73-1] (**186**), is an antihistamine marketed as the racemate. The (R)-(+)-isomer (**184**) is 65 times more potent than its (–)-enantiomer (**185**) when tested in guinea pig ileum (135). Chlorpheniramine (**187**) and (**188**) is also an enantioselective H₁-antagonist, wherein the (S)-enantiomer (**187**) is most potent (101). It has been demonstrated that the more potent enantiomer of diphenhydramine and pheniramine drugs is the one in which the aryl moiety, alkylamine group, and the *p*-substituted aryl functionality occur in a clockwise orientation (136).

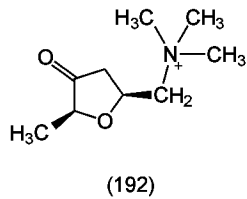
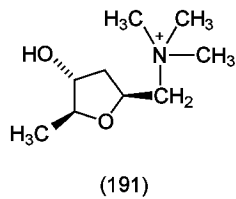




Acetyl choline is the natural neurotransmitter for the cholinergic receptor. Two distinct receptor subtypes have been characterized based on their binding affinity for either nicotine (189) and (190) or muscarine (191).

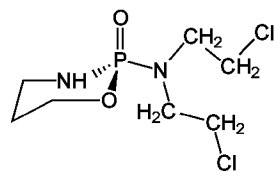


(S)-(-)-Nicotine [54-11-5] (189) is highly selective for the nicotinic receptor and is modestly more potent as an agonist than its (R)-(+)-isomer (190) (101). A 10-fold difference in binding density of ()- and (+)-nicotine ((S)-(-)-nicotine favored) in the rat brain P₂ fraction indicates that nicotine enantiomers may bind to different high affinity sites. Furthermore, (+)-nicotine enhances the binding of ()-nicotine at the ()-nicotine high affinity receptor site (137). The muscarinic receptor is highly stereoselective; (2*S*,3*R*,5*S*)-(+)-muscarine [300-54-9] (191) binds specifically to this receptor subtype whereas the other seven isomers are relatively inactive. Oxidation of (2*S*,3*R*,5*S*)-(+)-muscarine to (2*S*,5*S*)-muscarone (192) eliminates the stereoselectivity and receptor subtype selectivity observed for natural (+)-muscarine, while maintaining the cholinergic effects (138).

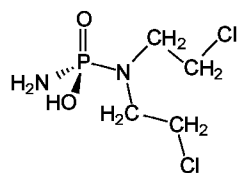


Antineoplastic Drugs. Cyclophosphamide (193) produces antineoplastic effects (see CHEMOTHERAPEUTICS,ANTICANCER) via biochemical conversion to a highly reactive phosphoramidate mustard (194); it is chiral owing to the tetrahedral phosphorus atom. The therapeutic index of the (S)-(-)-cyclophosphamide [50-18-0] (193) is twice that of the (+)-enantiomer due to increased antitumor activity; the enantiomers are equally toxic (139). The effectiveness of the DNA intercalator drugs adriamycin [57-22-7] (195) and daunomycin [20830-81-3] (196) is affected by changes in stereochemistry within the aglycon portions of these compounds. Inversion of the carbohydrate C-1 stereocenter provides compounds without activity. The carbohydrate C-4 epimer of adriamycin, epirubicin [56420-45-2], is as potent as its parent molecule, but is significantly less toxic (139).

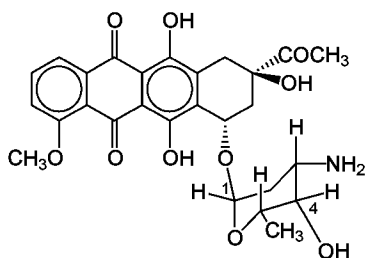
(R)-3-Ethyl-3(4-pyridyl)piperidine-2,6-dione (197), useful in the treatment of certain breast cancers, is a 20-fold more potent aromatase inhibitor ($IC_{50} = 10 \mu M$) than is its (*S*)-enantiomer (140).



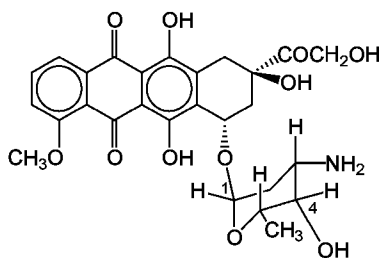
(193)



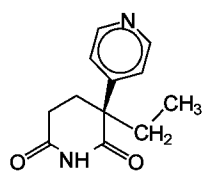
(194)



(195)



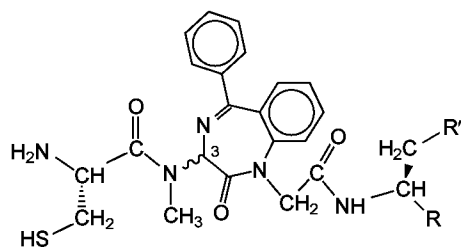
(196)



(197)

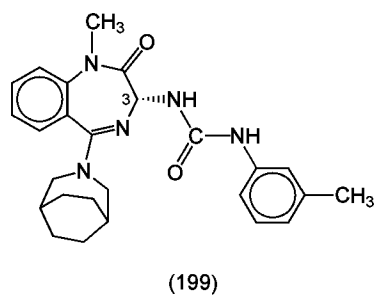
Peptidomimetics. Many drugs mimic natural small peptides. For example, morphine (168) is believed to be a natural peptidomimetic for the enkephalins. Similarly, FK-506 (105) mimics the binding of peptidal FK-506 to the intracellular receptor, FKBP12 (141). Numerous small endogenous peptides have been characterized which possess potent cellular signaling and homeostatic regulating activities. The regulation of glycolysis, growth, mitosis, and apoptosis, as well as the maintenance of blood pressure and the natural relief of pain, exemplify a few of the regulatory actions of peptide hormones (qv). Exogenous control, through the use of synthetic compounds, of the activities of these regulatory elements is highly desirable as demonstrated by the use of drugs such as the ACE inhibitor captopril (47). The design of peptidomimetics is complicated owing to the flexibility and stereochemical complexity of such hormones. Numerous small peptides have been synthesized which possess tremendous enzyme inhibitory and receptor binding activities *in vitro*; HIV protease inhibitors are one example. Unfortunately, the use of such peptides *in vivo* generally is not successful, as peptidase enzymes rapidly degrade synthetic peptides.

Several methods are being studied to enhance the stability of peptide mimics and improve their stereochemical similarity to the endogenous peptides. For example, the tetrapeptide Cys–Val–Phe–Met, a potent inhibitor of *Ras* farnesyltransferase, is proposed to exist in a turned conformation, which mimics the endogenous peptide during enzyme binding. This conformation is successfully mimicked by 3-amino-1-carboxymethyl-5-phenyl-benzodiazepin-2-one derivatives (198) (142).



(198)

Such benzodiazepine derivatives are potent inhibitors of the *Ras* farnesyltransferase enzyme. This, combined with their increased *in vivo* stability compared to peptide inhibitors, makes them good candidates for the treatment of *Ras* oncogene-dependent cancer. The benzodiazepine ring system has been successfully used in the development of other peptide mimics. Cholecystokinin (CCK) is a polypeptide hormone which occurs in numerous molecular forms throughout the peripheral and central nervous systems. CCK exerts a variety of actions on peripheral organs, such as regulating pancreatic secretion, gut motility, and gall bladder contraction. The actions of CCK are mediated by two receptor subtypes designated as CCK_A and CCK_B. The CCK receptor subtype selectivity of benzodiazepine L-740,093 (**199**) is regulated by the C3 stereochemistry of the benzodiazepine ring system.



L-740,093, possessing an (*R*) absolute stereochemistry, is highly selective for the CCK_B receptor (CCK_B / CCK_A = 16,000) with an IC₅₀ of 0.1 nM (143). The (*S*)-enantiomer demonstrates a fourfold selectivity for the CCK_A receptor with an IC₅₀ of 6.5 nM.

Use of D-amino acids in the synthesis of a hairpin loop portion from the CD4 receptor provides a stable CD4 receptor mimic, which blocks experimental allergic encephalomyelitis (144). This synthetic construct is not simply the mirror image or enantiomer of the CD4 hairpin loop, but rather an all-D-construct in the reverse sequence, thus providing stereochemically similar side-chain projections of the now inverted backbone (Fig. 11). This peptide mimetic, unlike its all-L amino acid counterpart, is resistant to enzyme degradation. As one would expect, the all-D amino acid CD4 hairpin loop, synthesized in the natural direction, the enantiomer of the natural construct, is inactive.

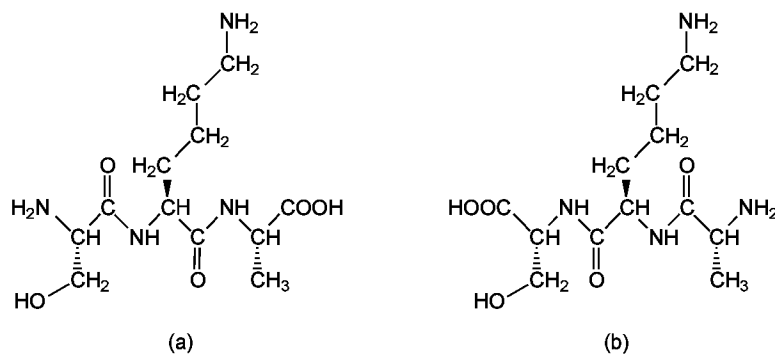


Fig. 11. Use of D-amino acids in the synthesis of a hairpin loop portion from the CD4 receptor: (a) all L-Ser-Lys-Ala tripeptide constructed in the natural direction; (b) all D-Ser-Lys-Ala tripeptide constructed in the reverse direction.

Economic Aspects of Homochiral Pharmaceuticals in Industry

Drugs classified as either natural or semisynthetic in origin accounted for ~22% of the market share in 1991 (145). Nearly 94% of the agents are chiral compounds and are sold as single enantiomers. Chiral synthetic drugs make up 38% of the market share and 43% of these are sold as single enantiomers, a two- to threefold increase since 1981. Achiral or symmetrical synthetic drugs make up 40% of the drug market. The vast number of marketed racemic drugs are being reinvestigated and newer pharmacological data as well as production technology are being patented (146).

The world sales of homochiral drugs grew 22% from 1992 to an estimated \$35.6 × 10⁶ in 1993. Enantiomeric cardiovascular drugs grossed \$11.3 × 10⁶ followed closely by antibiotics (\$10.8 × 10⁶), whereas hormones, CNS agents, antiinflammatory drugs, and antineoplastic drugs yielded a combined \$9 × 10⁶ (147). A steady increase in the number of homochiral drugs on world markets has created an increased demand for enantiomerically pure intermediates as well as for enantioselective technologies. Many pharmaceutical companies are pursuing new financial opportunities and gaining improved bargaining positions by producing patent protected and more expensive enantiomerically pure drugs from unprotected racemic pharmaceuticals.

BIBLIOGRAPHY

"Pharmaceuticals, Optically Active" in *ECT* 3rd ed., Vol. 17, pp. 311–345, by D. T. Witiak and M. N. Inbasekaran, Ohio State University.

1. J. Snopek and I. Jelinek, *J. Chromatograph.* **609**, 1–17 (1992).
2. P. R. Andrews, in H. Kubinyi, ed., *3D QSAR in Drug Design: Theory, Methods and Applications*, ESCOM Science Publishers, the Netherlands, 1993, pp. 13–21.
3. G. F. Russel and J. I. Hills, *Science* **172**, 1043–1044 (1971).
4. L. Freedman and J. G. Miller, *Science* **172**, 1044–1046 (1971).
5. J. Solms, L. Vuataz, and R. H. Egli, *Experientia* **21**, 692–694 (1965).
6. T. Z. Csaky, *Cutting's Handbook of Pharmacology*, 6th ed., Appleton-Century-Crofts, New York, 1979, p. 161.
7. H. Glatt and F. Oesch, *Biochem. Pharm.* **34**, 3725–3728 (1985).
8. G. von Blaschke, H. P. Kraft, K. Finkentscher, and F. Kohler, *Arzneim.-Forsch./Drug Res.* **29**, 1640 (1979).
9. K. Eger, M. Jalalian, E. J. Verspohl, and N.-P. Luke, *Arzneim. Forsch./Drug Res.* **40**, 1073–1075 (1990).
10. A. M. Evans, R. L. Nation, L. N. Sansom, F. Bochner, and A. A. Somogyi, *Br. J. Clin. Pharmac.* **26**, 771–780 (1988).
11. R. J. Ott and K. M. Giacomini, in I. W. Wainer, ed., *Drug Stereochemistry Analytical Methods and Pharmacology Second Edition, Revised and Expanded*, Marcel Dekker, Inc., New York, 1993, pp. 281–314.
12. T. A. G. Noctor, in Ref. 11, pp. 337–364.
13. N. P. E. Vermeulen and J. M. Koppele, in Ref. 11.
14. B. Holmstedt, in B. Holmstedt, H. Frank, and B. Testa, eds., *Chirality and Biological Activity*, Alan R. Liss, Inc., New York, 1990, pp. 1–14.
15. S. Hoffmann, in R. Janoschek, ed., *Chirality—From Weak Bosons to the α-Helix*, Springer-Verlag, Berlin, Heidelberg, 1991, p. 206.
16. S. Kotha, *Tetrahedron* **50**, 3639–3662 (1994).

17. S. Levin and S. Abu-Lafi, in P. R. Brown and E. Grushka, eds., *Advances in Chromatography*, Vol. 33, Marcel Dekker, Inc., New York, 1993, pp. 233–266.
18. E. J. Arians, *TiPS* **14**, 68–75 (1993).
19. E. J. Arians, E. W. Wuis, and E. J. Veringa, *Biochem. Pharmacol.* **37**, 9–18 (1988).
20. B. Holmstedt, H. Frank, and B. Testa, *Chirality and Biological Activity*, Alan R. Liss, Inc., New York, 1990.
21. U.S. Food and Drug Administration, *Chirality* **4**, 338–340 (1992).
22. W. H. De Camp, *Chirality* **1**, 2–6 (1989).
23. B. Testa, in Ref. 14, pp. 15–32.
24. W. L. Alworth, *Stereochemistry and Its Application in Biochemistry*, Wiley-Interscience, New York, 1972, pp. 60, 163–177.
25. G. Snatzke, *Optical Rotatory Dispersion and Circular Dichroism in Organic Chemistry*, Sadtler Research Laboratories, Inc., Philadelphia, Pa., 1967.
26. V. Prelog and G. Helmchen, *Angew. Chem. Int. Ed. Engl.* **21**, 567–583 (1982).
27. J. K. O’Loane, *Chem. Rev.* **80**, 41–61 (1980).
28. B. Testa, in C. Tamm, ed., *Stereochemistry*, Elsevier Biomedical Press, Amsterdam, the Netherlands, 1982, pp. 1–47.
29. R. Janoschek, *Chirality—From Weak Bosons to the α -Helix*, Springer-Verlag, Berlin, Heidelberg, 1991.
30. M. Hyneck, J. Dent, and J. B. Hook, in C. Brown, ed., *Chirality in Drug Design and Synthesis*, Academic Press, Inc., San Diego, Calif., 1990, pp. 1–28.
31. E. J. Arians, in Ref. 30, pp. 29–51.
32. C. R. Ganellin and S. M. Roberts, eds., *Medicinal Chemistry, The Role of Organic Chemistry in Drug Research*, 2nd ed., Academic Press, Inc., San Diego, Calif., 1993.
33. S. Mason, *New Sci.* **165**, 10–14 (1984).
34. S. Mason, *TiPS* **7**, 20–23 (1986).
35. J. L. Bada and S. L. Miller, *Biosystems* **20**, 21–26 (1987).
36. R. Janoschek, in Ref. 29, pp. 18–33.
37. M. Friedman and M. R. Grumbmann, in M. Friedman, ed., *Absorption and Utilization of Amino Acids*, Vol. II, CRC Press, Boca Raton, Fla., 1989, pp. 173–190.
38. E. H. Man and J. L. Bada, *Ann. Rev. Nutr.* **7**, 209–225 (1987).
39. E. G. Maliski and J. Bradshaw, in Ref. 32, pp. 83–102.
40. A. Fersht, *Enzyme Structure and Mechanism*, W. H. Freeman and Co., New York, 1985.
41. H. Kubinyi, in Ref. 2.
42. U. Hacksell, A. M. Johansson, A. Karlen, K. Luthman, and C. Mellin, in Ref. 30, pp. 83–100.
43. S. Wang and co-workers, *J. Med. Chem.* **37**, 1326–1338 (1994).
44. S. M. Roberts, in Ref. 30, pp. 115–214.
45. J. W. Scott, in Ref. 11, pp. 183–212.
46. J. Jacques, A. Collet, and S. H. Wilen, *Enantiomers, Racemates and Resolutions*, Wiley-Interscience, New York, 1981.
47. K. H. Bui, in G. W. Fong and S. K. Lam, eds., *HPLC in the Pharmaceutical Industry*, Vol. 47, Marcel Dekker, Inc., New York, 1990, pp. 213–239.
48. A. M. Kristulovic, *Chiral Separations by HPLC*, Ellis Horwood, Chichester, U.K., 1989.
49. K. Imai, in J. C. Giddings, E. Grushka, and P. R. Brown, eds., *Advances in Chromatography*, Vol. 27, Marcel Dekker, Inc., New York, 1987, pp. 215–245.
50. W. Lindner, J. N. LePage, G. Davies, D. E. Seitz, and B. L. Karger, *J. Chromatogr.* **185**, 323–344 (1979).
51. P. E. Hare, E. Gil-Av, *Science* **204**, 1226–1228 (1979).
52. V. A. Davankov, A. A. Kurganov, and A. S. Bochkov, in J. C. Giddings, E. Grushka, J. Cazes, and P. R. Brown, eds., *Advances in Chromatography*, Vol. 22, Marcel Dekker, New York, 1983, pp. 71–116.
53. W. H. Pirkle and T. C. Pochapsky, in Ref. 49, pp. 73–127.
54. H. Y. Aboul-Enein, in S. Ahuja, ed., *Chromatography of Pharmaceuticals Natural, Synthetic, and Recombinant Products*, American Chemical Society, Washington, D.C., 1992, pp. 111–120.
55. G. Gubitz, *Chromatographia* **30**, 555–564 (1990).
56. D. R. Taylor and K. Maher, *J. Chrom. Sci.* **30**, 67–85 (1992).
57. H. G. Davies, R. H. Green, D. R. Kelly, and S. M. Roberts, *Bio-Transformations in Preparative Organic Chemistry: The Use of Isolated Enzymes and Whole-Cell Systems in Synthesis*, Academic Press, London, 1989.
58. K. Faber, *Bio-Transformations in Organic Chemistry*, Springer-Verlag, New York, 1992, pp. 47–48.
59. A. M. Klibanov, *Acc. Chem. Res.* **23**, 114–120 (1990).
60. M. Ohno and M. Otsuka, *Org. Reac.* **37**, 1–55 (1989).
61. Y. Terao, M. Akamatsu, and K. Achiwa, *Chem. Pharm. Bull.* **39**, 823–825 (1991).
62. D. I. Stirling, in Ref. 11, pp. 213–243.
63. R. A. Sheldon, *Chirality Technology Industrial Synthesis of Optically Active Compounds*, Marcel Dekker, Inc., New York, 1993.
64. Q. M. Gu, C. S. Chen, and C. J. Sih, *Tetrahedron Lett.* **27**, 1763 (1986).
65. R. A. Sheldon, *Chirality Technology Industrial Synthesis of Optically Active Compounds*, Marcel Dekker, Inc., New York, 1993, pp. 243, 362–367.
66. R. A. Sheldon, in R. A. Sheldon, ed., *Chirality Technology Industrial Synthesis of Optically Active Compounds*, Marcel Dekker, Inc., New York, 1993, pp. 205–270.
67. G. M. Coppola and H. F. Schuster, *Asymmetric Synthesis: Construction of Chiral Molecules Using Amino Acids*, John Wiley & Sons, Inc., New York, 1987.
68. J. D. Morrison and J. W. Scott, *Asymmetric Synthesis*, Vol. 4, *The Chiral Carbon Pool and Chiral Sulfur, Nitrogen, Phosphorus, and Silicon Centers*, Academic Press, Inc., Orlando, Fla., 1984.
69. R. A. Sheldon, in Ref. 66, pp. 143–171.
70. D. B. Collum, J. H. McDonald, III, and W. C. Still, *J. Am. Chem. Soc.* **102**, 2117–2118 (1980).
71. *Ibid.*, pp. 2118–2120.
72. *Ibid.*, pp. 2120–2121.
73. R. A. Holton and co-workers, *J. Am. Chem. Soc.* **116**, 1597–1598 (1994).
74. D. A. Evans, H. P. Ng, and D. L. Rieger, *J. Am. Chem. Soc.* **115**, 11446–11459 (1993).
75. W. Oppolzer, in Ref. 30, pp. 199–214.
76. J. D. Morrison, *Asymmetric Synthesis*, Vols. 1–5, Academic Press, Inc., San Diego, Calif., 1983.
77. M. M. Midland and A. Tramontano, *Tetrahedron Lett.* **21**, 3549–3552 (1980).
78. A. T. Hopper and D. T. Witiak, *J. Org. Chem.* **60**, 3334–3341 (1995).
79. K. B. Sharpless, *Chemtech*, 692–700 (1985).
80. P. Kocienski, M. Stocks, and D. K. Donald, in Ref. 30, pp. 131–165.
81. H. B. Kagan, in W. Bartmann and K. B. Sharpless, eds., *Stereochemistry of Organic and Bioorganic Transformation*, Verlagsgesellschaft mbH, Weinheim, Germany, 1987, pp. 31–45.

82. R. A. Sheldon, *Chirotechnolgy: Industrial Synthesis of Optically Active Compounds*, Marcel Dekker, Inc., New York, 1993, pp. 271–341.
83. R. Noyori, *Chemtech*, 360–367 (1992).
84. M. Kitamura, Y. Hsiao, R. Noyori, and H. Takaya, *Tetrahedron Lett.* **28**, 4829–4832 (1987).
85. C. Kratky, in Ref. 29, pp. 86–102.
86. P. S. Kalsi, *Stereo-Chemistry Conformation and Mechanism*, John Wiley & Sons, New Delhi, India, 1990, pp. 277–283.
87. Ref. 86, pp. 284–290.
88. J. Caldwell and B. Testa, *Drug Metab. Dispos.* **15**, 587–588 (1987).
89. G. G. Lyle and R. E. Lyle, in J. D. Morrison, ed., *Asymmetric Synthesis*, Vol. 1, *Analytical Methods*, Academic Press, Inc., San Diego, Calif., 1983, pp. 13–27.
90. R. R. Fraser, in Ref. 89, pp. 173–196.
91. J. A. Dale, D. L. Dull, and H. S. Mosher, *J. Org. Chem.* **34**, 2543–2549 (1969).
92. S. Yamaguchi, in Ref. 89, pp. 125–152.
93. D. F. Smith, *CRC Handbook of Stereoisomers: Therapeutic Drugs*, CRC Press, Inc., Boca Raton, Fla., 1989.
94. B. G. Main and H. Tucker, in Ref. 32, pp. 187–208.
95. C. K. Buckner and P. N. Patil, *J. Pharm. Exp. Ther.* **176**, 634–649 (1971).
96. A. J. Coleman, D. S. Paterson, and A. R. Somerville, *Biochem. Pharmacol.* **28**, 1011–1013 (1979).
97. M. J. Mardle, H. Smith, and B. A. Spicer, *J. Med. Chem.* **17**, 513–516 (1974).
98. W. J. Janssens, R. Xhonneux, and P. A. J. Janssen, *Drug Inv.* **3**(Suppl. 1), 13–24 (1991).
99. P. N. Patil, J. B. LaPidus, and A. Tye, *J. Pharmacol. Exp. Ther.* **155**, 1–12 (1967).
100. R. R. Ruffolo, in E. J. Arians, W. Soudijn, and P. B. M. W. M. Timmermans, eds., *Stereochemistry and Biological Activity of Drugs*, Blackwell Scientific, Oxford, U.K., 1983, pp. 103–125.
101. M. Simonyi, *Med. Res. Rev.* **4**, 359–413 (1984).
102. P. B. M. W. M. Timmermans and M. J. M. C. Thoolen, in Ref. 93, pp. 13–33.
103. R. R. Ruffolo, E. L. Yaden, and J. E. Waddell, *J. Pharmacol. Exp. Ther.* **222**, 645–651 (1982).
104. L. H. Easson and E. Stedman, *Biochem. J.* **27**, 1257–1266 (1933).
105. A. H. Beckett, *Fortschr. Arzneim. Forsch.* **1**, 455–530 (1959).
106. D. G. Berry and J. W. Miller, *Eur. J. Pharm.* **28**, 164–171 (1974).
107. G. Stemp and J. M. Evans, in Ref. 32, pp. 141–162.
108. V. A. Ashwood and co-workers, *J. Med. Chem.* **29**, 2194–2201 (1986).
109. M. A. Abou-Gharbia, D. H. Kim, and R. J. McCaully, in Ref. 93, pp. 65–124.
110. C. S. Sweet, A. A. Patchett, E. H. Ulm, D. M. Gross, in J. C. Emmett, ed., *Second SCI-RSC Medicinal Chemistry Symposium*, The Royal Society of Chemistry, Burlington House, London, 1984, pp. 36–46.
111. R. A. Janis and D. J. Triggle, *J. Med. Chem.* **26**, 775–785 (1983).
112. D. J. Triggle, in Ref. 110, pp. 1–17.
113. J. A. Vida, in W. O. Foye, ed., *Principles of Medicinal Chemistry*, 3rd ed., Lea and Febiger, Philadelphia, Pa., 1989, pp. 143–171.
114. D. F. Smith, in Ref. 93, pp. 323–393.
115. W. Soudijn, in Ref. 100, pp. 91–92.
116. C. E. Cook, T. B. Seltzman, C. R. Tallent, B. Lorenzo, and D. E. Drayer, *J. Pharmacol. Exp. Ther.* **241**, 779–785 (1987).
117. A. S. Nies, G. H. Evans, and D. G. Shand, *J. Pharmacol. Exp. Ther.* **184**, 716–720 (1973).
118. A. Corbella and co-workers, *J. Chem. Soc. Chem. Commun.* **19**, 721–722 (1973).
119. L. A. Mitscher, G. I. Georg, N. Motohashi, in Ref. 93, pp. 199–234.
120. J. F. Gerster, *J. Med. Chem.* **30**, 839–843 (1987).
121. L. A. Mitscher, P. N. Sharma, D. T. W. Chu, L. L. Shen, and A. G. Pernet, *J. Med. Chem.* **30**, 2283–2286 (1987).
122. O. Okazaki, C. Kojima, H. Hakusui, and M. Nakashima, *Antimicrob. Agents Chemother.* **35**, 2106–2109 (1991).
123. I. Tamai, H.-Y. Ling, S.-M. Timbul, J. Nishikido, and A. Tsuji, *J. Pharm. Pharmacol.* **40**, 320–324 (1988).
124. A. F. Casy, in Ref. 93, pp. 285–321.
125. D. E. D. Drayer, *Clin. Pharm. Ther.* **40**, 125–133 (1986).
126. E. L. May, *J. Med. Chem.* **35**, 3587–3594 (1992).
127. F. Jamali, R. Mehvar, and F. M. Pasutto, *J. Pharm. Sci.* **78**, 695–715 (1986).
128. B. R. Martin and co-workers, *J. Pharmacol. Exp. Ther.* **231**, 539–544 (1984).
129. C. J. Evans, D. L. Hammond, R. C. A. Frederickson, in G. W. Pasternak, ed., *The Opioid Receptors*, The Humana Press, Clifton, N.J., 1988, pp. 23–71.
130. A. Z. Rónai, in J. I. Szakely and A. Z. Rónai, eds., *Opioid Peptides*, Vol. III, *Opiate Receptors and their Ligands*, CRC Press, Boca Raton, Fla, 1982, pp. 141–175.
131. G. Feuerstein, in R. S. Rapaka, G. Barnett, and R. L. Hawks, eds., *Opioid Peptides: Medicinal Chemistry*, National Institute on Drug Abuse Research Monograph Series, Vol. 69, 1986, Rockville, Md., pp. 112–127.
132. R. A. O'Reilly and W. F. Trager, in Ref. 95, pp. 125–136.
133. L. G. M. Baars, M. T. Schepers, J. J. R. Hermans, H. J. J. Dahlmans, and H. H. W. Thijssen, *J. Pharm. Pharmacol.* **42**, 861–866 (1990).
134. C. Banfield, R. O'Reilly, E. Chan, and M. Rowland, *Br. J. Clin Pharmacol.* **16**, 669–675 (1983).
135. A. F. Casy, in Ref. 93, pp. 149–164.
136. A. F. Casy, A. F. Drake, C. R. Ganellin, A. D. Mercer, and C. Upton, *Chirality*, **4**, 356–366 (1992).
137. J. W. Sloan, W. R. Martin, J. Hernandez, and R. Hook, *Pharmacol. Biochem. Behav.* **23**, 987–993 (1985).
138. R. Dahlbom, in Ref. 100, pp. 127–142.
139. E. von Angerer, E. Holler, H. Schonenberger, and R. Schonenberger, in Ref. 93, pp. 247–284.
140. R. McCague and co-workers, *J. Chem. Soc., Perkin Trans. I.*, 196–200 (1989).
141. Y. Ikeda, L. W. Schultz, J. Clardy, and S. L. Schreiber, *J. Am. Chem. Soc.* **116**, 4143–4144 (1994).
142. G. L. James and co-workers, *Science* **260**, 1937–1942 (1993).
143. G. A. Showell and co-workers, *J. Med. Chem.* **37**, 719–721 (1994).
144. B. A. Jameson, J. M. McDonnell, J. C. Marini, and R. Korngold, *Nature*, **368**, 744–746 (1994).
145. J. S. Millership and A. Fitzpatrick, *Chirality*, **5**, 573–576 (1993).
146. S. C. Stinson, *Chem. Eng. News*, 38–68 (Sept. 27, 1993).
147. S. C. Stinson, *Chem. Eng. News*, 38–72 (Sept. 19, 1994).

Donald T. Witiak

Allen T. Hopper
University of Wisconsin-Madison

PHARMACEUTICALS, CONTROLLED RELEASE.

See CONTROLLED RELEASE TECHNOLOGY; DRUG DELIVERY SYSTEMS.

PHARMACODYNAMICS

Pharmacology relates to all aspects of drug action, including drug synthesis or isolation from natural products, elucidation of physiologic and toxicologic effects, determination of therapeutic applications, and mechanisms of action (1,2). Drugs, chemical species that affect cellular function with specificity, include hormones (qv), neurotransmitters (see NEUROREGULATORS; OPIOIDS, ENDOGENOUS), and chemotherapeutic agents (see ANTI-ASTHMATIC AGENTS; ANTIBIOTICS; ANTIPARASITIC AGENTS; CHEMOTHERAPEUTICS, ANTICANCER). The principles of chemical specificity are common to all of these classes of agents. Pharmacodynamics is the study of drug action primarily in terms of drug structure, site of action, and the biochemical and physiological consequences of the drug action. The availability of a drug at its site of action is determined by several processes (Fig. 1), including absorption, metabolism, distribution, and excretion. These processes constitute the pharmacokinetic aspects of drug action. The onset, intensity, and duration of drug action are determined by these factors as well as by the availability of the drug at its receptor site(s) and the events initiated by receptor activation (see DRUG DELIVERY).

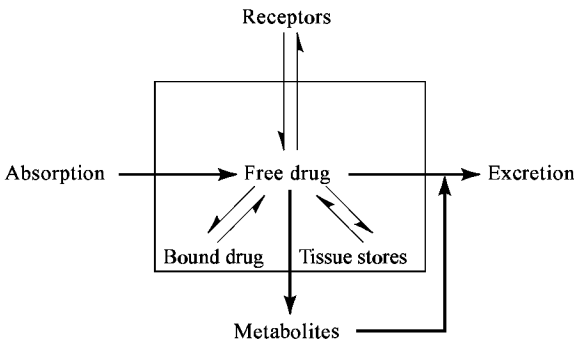


Fig. 1. Schematic representation of drug disposition following administration.

Both pharmacokinetic and pharmacodynamic processes are involved in mediating nonconstant expressions of drug action. Thus, resistance to the actions of a drug, eg, in the development of antibiotic-resistant bacteria or of barbiturate tolerance, can arise from changes in drug metabolism and or alterations in the receptor target site. Factors controlling drug resistance may be whole-body, cellular, or individual events. Decreased absorption, increased metabolism, or increased elimination reduce circulating drug levels and affect the whole body. Increased drug metabolism, increased concentration of an agent that antagonizes drug action, decreased affinity or concentration of a drug receptor, and depletion of an agent that mediates drug action are examples of cellular events; and genetic factors controlling metabolism, receptor alterations, and disease states are examples of individual events (1). Individual variation in the susceptibility to a particular drug or class of drugs also may arise from genetically based pharmacokinetic factors as well as from specific receptor-linked changes. For a large number of drugs, including neurotransmitters, peptide and protein hormones (qv), and their analogues and antagonists, the cell membrane is the principal locus of action. Concepts of cell membrane structure are derived from the original Davson-Danielli lipid bilayer hypothesis. More specifically, the membrane is viewed as a dynamic fluid mosaic or a matrix of fluid bilayer in which there are asymmetrically inserted proteins (qv) and glycoproteins. Phospholipids and proteins diffuse laterally and the resultant protein–protein communication is of considerable importance to the understanding of membrane-receptor function (3). Despite the dynamic nature of the membrane and the absence of global organization, local organization is possible through the local assembly of individual protein components and the attachment of membrane proteins to the subcellular structure of contractile proteins. However, the cell membrane is not the site of action of all drugs. A number of drugs, including steroid and thyroid hormones, exert their effects intracellularly at the level of the genetic material as well as at the plasma membrane (see STEROIDS; THYROID AND ANTITHYROID PREPARATIONS). Other agents, including polypeptide growth factors, exert their effects not only at the plasma membrane through tyrosine kinase receptors, but also on cell growth and differentiation at the genetic level.

Drug Discovery and Regulation

The Federal Food, Drug and Cosmetic (FDAC) Act defines drugs as "...articles intended for use in the diagnosis, cure, mitigation, treatment, or prevention of disease in man..." and "articles (other than food) intended to affect the structure of any function of the body of man." In the United States and elsewhere, the introduction of a new drug is subject to a sequence of well-defined stages of development and approval (4). Each stage involves either scientific testing or submission and preparation of data and analysis review (Fig. 2).

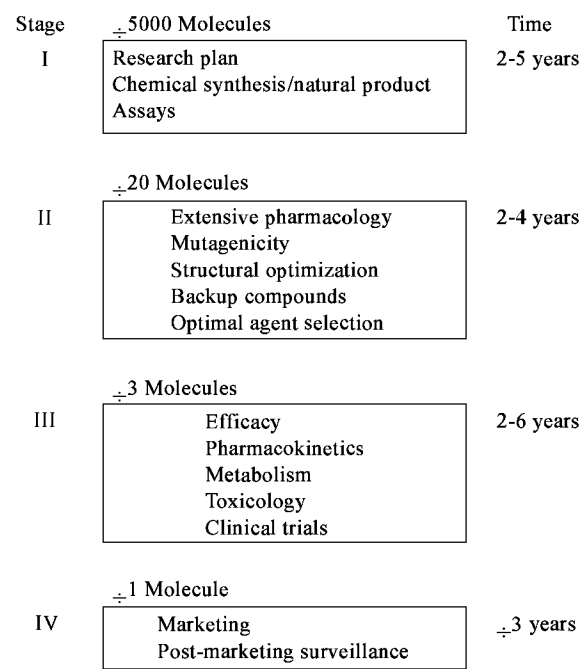


Fig. 2. Pathway for drug development.

An investigational new drug (IND) application usually initiates the process for drug approval. The IND derives from the concept that a specific molecule or molecules may have a particular therapeutic benefit. Preclinical data are analyzed to determine the implications of such molecules for human pharmacology, chemical composition, manufacturing processes, and the protocols for subsequent clinical work. Clinical trials are usually carried out in at least three phases. Phase one involves a small number of individuals and is designed to find information about basic safety and response issues. In phase two studies, the drug is employed on a larger number of individuals (100–200) who suffer from the condition that the drug is designed to treat. Phase three studies involve a much larger group of patients and are designed to assess safety, efficacy, and dosage regimens in a broad range of patients across lines of age, race, and gender. Phase three studies may involve several thousand patients and be carried out at several sites.

New drug application (NDA) is the process through which the U.S. Food and Drug Administration (FDA) authorizes the marketing of a new drug. In the NDA, the data are intended to demonstrate the safety and efficacy of the drug in its intended application. After approval, the drug becomes available to the public. Subsequently, dosage amounts and forms may be modified according to experience, new indications may be added, and contraindications may be noted. All of the changes require regulatory approval. A drug in human use is subject to constant surveillance.

The Receptor Concept

The concept of drug–receptor interaction originates in two separate lines of experimental evidence advanced in the late nineteenth and early twentieth centuries (5,6). Work in immunology and the chemotherapy of protozoan infections led to the postulate that specific protoplasmic side chains of unique chemical and steric architecture exist, and these side chains combine in lock-and-key fashion only with the appropriate complementary groups of an antibody or a chemotherapeutic agent (7). At about the same time, the concept of a receptive substance was developed, based on the mutually antagonistic actions of atropine and pilocarpine (8). The former inhibits saliva flow; the latter stimulates it. Subsequently, on the basis of the antagonist effects of curare on nicotine-induced skeletal muscle contraction, it was concluded that a specially excitable constituent, ie, the receptor, exists (9). Thus, early attention was drawn both to the specific recognition capacity of receptors and to the ability of a drug–receptor complex to initiate a biological response. The principle of specific chemical recognition is common to ligand–macromolecule interactions, but this alone does not suffice to define a receptor in the pharmacologic sense. Rather, it is the combination of chemical specificity or recognition and the capacity to initiate biological response or transduction that define the pharmacologic receptor (1,10,11).

The information contained in the chemical structure of a given ligand is without value unless decoded and executed by the appropriate receptor. The pharmacologic analysis of drug–receptor interactions is based on the understanding of how the drug is recognized by the receptor, how the drug–receptor complex forms, and how the drug–receptor complex initiates its biological action (12).

Drug receptors are chemical entities which are typically, but not exclusively, small molecules that interact with cellular components, frequently at the plasma membrane level (1,2). There are many types of receptors; heat, light, immune, hormone, ion channel, toxin, and virus are but a few that can excite a cell. The receptor concept can be applied generally to signal recognition processes where a chemical or physical signal is recognized. This recognition is translated into response (Fig. 3) and the process can be seen as a flow of information.

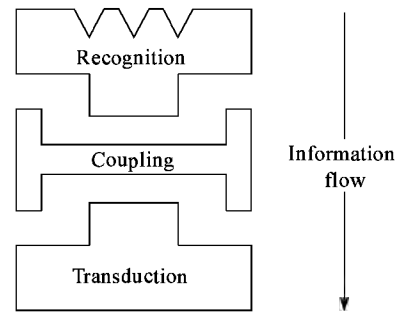


Fig. 3. Information flow at a pharmacologic receptor. The receptor is depicted as having the three components recognition, coupling, and transduction. These may be separate components or carried on a single function.

Elucidation of the structural requirements for drug interaction at the recognition site is by the study of structure–activity relationships (SAR), in which, according to a specific biologic response, the effects of systematic molecular modification of a parent drug structure are determined. Such studies have permitted the classification of discrete classes of pharmacological receptors. For example, the neurotransmitter acetylcholine acts at both peripheral and central receptors which are of at least three distinct types. The effects of acetylcholine are mimicked in smooth and cardiac muscles and secretory

glands by muscarine and in skeletal muscle and autonomic ganglia by nicotine; the effects in smooth and cardiac muscles are antagonized by atropine, in skeletal muscle by curare, and in ganglia by hexamethonium. Thus, it is recognized that there are at least two primary classes of agents acting at receptors: those that bind to the receptor and initiate its specific response, ie, agonists, and those that bind but are unable to initiate response, ie, antagonists. Additionally, selective chemical agents indicate the existence of subclasses of a primary acetylcholine-recognizing subunit.

The demonstration of the existence of strictly defined SARs, which is perhaps the most important criterion of drug action at a specific receptor site, has made possible the most important pharmacologic discoveries. For example, the analgesic actions of morphine [57-27-2] and related agents, which are indicative of specific receptors, led to the discovery of endogenous opiate peptides, ie, the leucine and methionine enkephalins and endorphins (see OPIOIDS, ENDOGENOUS) (13). Similarly, the well-defined SAR for the anxiolytic activity of the benzodiazepines, eg, diazepam [439-14-5] (Valium) and flurazepam [17617-23-1], also indicates that endogenous compounds which interact in a physiological fashion with these receptors exist (see PSYCHOPHARMACOLOGICAL AGENTS; STIMULANTS) (14). In the 1990s, the existence of SARs for cannabinoids (Marihuana) led first to the discovery of the cannabinoid receptor, which, like the opiate receptor, is also a plasma membrane G-protein-coupled protein, and then to the discovery of an endogenous cannabinoid-receptor active factor (or factors) that may serve to mediate physiological events (15–17).

Pharmacokinetic Aspects of Drug Action

The receptor represents the locus of drug action. However, the pharmacokinetic processes of absorption (drug entry), distribution, metabolism, and excretion play principal roles in determining *in vivo* time courses and concentrations of drugs and thus modify actions initiated at receptors.

Drug Entry. Drugs enter the body by one of two routes. In enteral administration (sublingual, oral, rectal), the drug enters directly the gastrointestinal tract. In the parenteral route, the drug bypasses the gastrointestinal tract by, among others, subcutaneous (sc), intramuscular, intravascular (iv), inhalational, intraperitoneal (ip), intravaginal, and intranasal routes. Each route has a particular set of advantages and disadvantages. Patient convenience is high in the oral route; speed of action and ability to control concentrations are high in the iv route; and nonoral routes are best for unstable or insoluble drugs.

In light of the recognized importance of achieving stable, reproducible plasma concentrations of drugs, particular attention is given to pathways and devices, including sustained-release formulations, pumps, and transdermal entry processes that ensure such properties (see CONTROLLED RELEASE TECHNOLOGY, PHARMACEUTICALS). Osmotic pumps are available for subcutaneous implants and for oral administration. These capsule-like devices contain a semipermeable membrane through which water can enter to dissolve the drug or to push the drug solution out of the system (1).

Drug Distribution. After administration, a drug may be distributed either generally or selectively in the body. The distribution pattern depends on many factors, including the pattern and time-course of blood flow, diffusion of drugs into tissues, binding of drugs to plasma proteins and cellular compartments, and elimination kinetics and mechanisms. The total body water of an average, 70-kg individual is around 42 L, which consists of 14 L of extracellular fluid which in turn includes about 3 L of plasma water. The apparent volume of distribution represents the volume of fluid in which a drug appears to be dissolved. A volume of distribution approximating that of the extracellular water indicates the locus for the drug, whereas a volume of distribution significantly higher or lower indicates a drug distribution into additional or restricted compartments, respectively (18).

Many drugs bind to plasma proteins. Such binding affects distribution and access to sites of action, metabolism, and elimination. Some drugs interact with specific plasma proteins, including the α - and β -lipoproteins for vitamins A and other carotenoids as well as the steroid hormones, the sex-steroid binding protein, and transcortin. A significant contribution to drug binding is made by albumin (mol wt ca 68,000), which constitutes some 50% of the total protein of plasma. Serum albumin contains multiple drug binding sites and is the endogenous carrier for free fatty acids. When drugs are extensively bound to albumin, competition between different drugs may underlie some drug interactions. This is well established for such drugs as warfarin [81-81-2] and the calcium antagonists, eg, verapamil [52-53-9], nifedipine [21829-25-4], and diltiazem [42399-41-7] (see CARDIOVASCULAR AGENTS). Although albumin has a high affinity for acidic drugs, basic species tend to bind to other proteins, including α_1 -acid glycoprotein, present at levels 50–100 times lower than albumin (19). α_1 -acid glycoprotein is an acute-phase reactant protein and its plasma levels are subject to great variation according to physiologic and pathologic conditions. These changes influence drug binding and the expression of drug activities.

Drug distribution into tissue reservoirs depends on the physicochemical properties of the drug. Tissue reservoirs include fat, bone, and the principal body organs. Access of drugs to these reservoirs depends on partition coefficient, charge or degree of ionization at physiological pH, and extent of protein binding. Thus, lipophilic molecules accumulate in fat reservoirs and this accumulation can alter considerably both the duration and the concentration–response curves of drug action. Some drugs may accumulate selectively in defined tissues, for example, the tetracycline antibiotics in bone (see ANTIBIOTICS, TETRACYCLINES).

Specific barriers may serve to limit drug distribution. The placental barrier is of obvious importance to drug action in the fetus. Drug transfers across the placenta primarily by lipid solubility. Hence, this barrier is not particularly restrictive. Similarly, the lipid solubility of a drug is a primary determinant in access to the brain and cerebrospinal fluid. Generally, hydrophilic or charged drugs can also penetrate to these latter areas, but the result is slow and incomplete. The blood brain barrier is composed of cells having tight junctions which are much less permeable to solutes than are the endothelial cells of other tissues.

Drugs may also accumulate selectively in reservoirs by active processes. A number of agents, including catecholamines, choline [62-49-7] (qv), and amino acids (qv) such as glutamate and γ -aminobutyric acid [56-12-2] (GABA), are taken up into cells via a Na^+ -dependent countertransport system (20). This system derives from the ion gradients established by the Na^+ - and K^+ -adenosinetriphosphotase (ATPase) pump. These transporters are the sites for many important therapeutic agents, including the antidepressants imipramine [50-49-7] and fluoxetine [54910-89-3], as well as drugs of abuse (see STIMULANTS). Drugs may also enter cells by variations on the theme of receptor-mediated endocytosis. Thus low density lipoprotein (LDL) enters cells by initial complexation to the LDL-receptor. Internalization via a specific coated-pit pathway is a key process in the metabolic control of cholesterol [57-88-5] biosynthesis and the regulation of plasma LDL receptors. This pathway and other internalization pathways are also participants in various down-regulation processes of receptors during chronic drug treatment and disease states.

Drug Metabolism. Generally, metabolism (biotransformation) of drugs increases their water solubility as well as the rate and ease of elimination, but reduces their volume of distribution. Many drug-metabolizing pathways have arisen during evolution to deal with foreign compounds present in food materials. Although metabolism generally leads to more polar and less active compounds, there are exceptions. Metabolic pathways have also been exploited to design prodrugs, materials that are converted to active species through biotransformation (1,2).

Biotransformation reactions can be classified as phase I and phase II. In phase I reactions, drugs are converted to product by processes of functionalization, including oxidation, reduction, dealkylation, and hydrolysis. Phase II or synthetic reactions involve coupling the drug or its polar metabolite to endogenous substrates and include methylation, acetylation, and glucuronidation (Table 1).

Table 1. Biotransformation Reactions

Pathways		Reactions types	Examples
		<i>Phase I reactions</i>	
oxidative		aliphatic and aromatic oxidation	phenobarbital, phenytoin
		N- and O-dealkylation	desipramine, phenacetin
		N-oxidation	guanethidine
		oxidative deamination	amphetamine
		desulfuration	thiobarbitol
		dehalogenation	chloroform
hydrolytic		esters and amides	procaine, lidocaine
reductive		azo reduction	prontosil

nonmicrosomal oxidative	nitro reduction	chloramphenicol
	alcohol and aldehyde oxidation	ethanol
	purine oxidation	6-mercaptopurine
	oxidative deamination	serotonin
	(monoamine oxidase)	
coupling	<i>Phase II reactions</i>	
	glucuronidation	acetaminophen
	acetylation	isoniazid
	glycine conjugation	salicylic acid
	sulfate conjugation	steroids, phenols
	methylation	norepinephrine

The liver microsomal drug-metabolizing system is of particular importance. This oxidative pathway is mediated by isozymes of the cytochrome P₄₅₀ family (Fig. 4). At least ten P₄₅₀ enzyme families exist to accommodate the ability of humans to handle many foreign molecules (21).

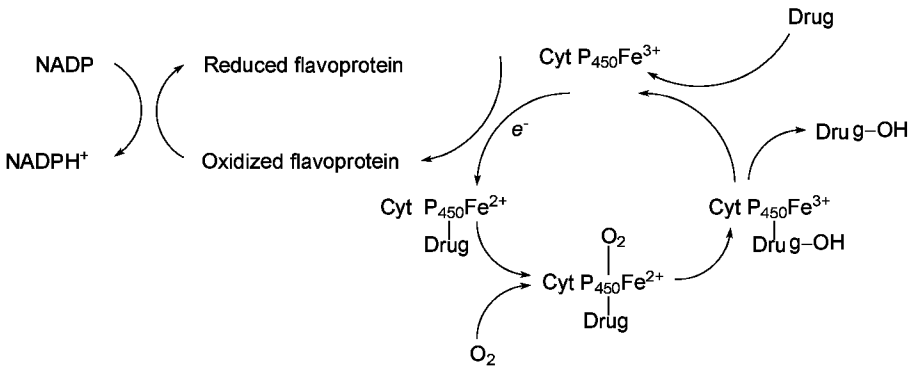
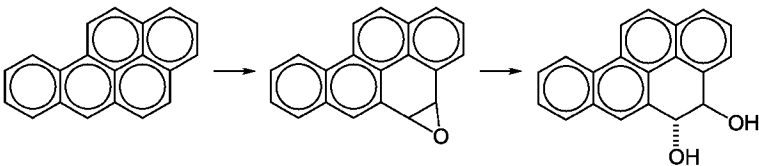


Fig. 4. The cycle of events involved in cytochrome (Cyt) P₄₅₀ (Cyt P₄₅₀ Fe²⁺) mediated drug metabolism where NADP is nicotinamide diphosphate and NADPH⁺ is the reduced form of NADP.

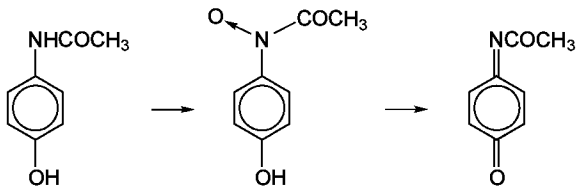
The biotransforming pathways are subject to manipulation and modification in a variety of ways. Drug metabolism can be induced as in hepatic enzyme induction, whereby a variety of agents, including barbiturates, aromatic hydrocarbons, and steroids, actually increase the amount of P₄₅₀ enzymes and thus enhance their own metabolism as well as that of other substrates. A number of agents serve as inhibitors of drug metabolism, including SKF525A (2-diethylamino-diphenylpropylacetate), the histamine H₂ antagonists cimetidine [51481-61-9] and disulfiram [97-77-8] (Antabuse). Thus, cimetidine can inhibit metabolism, potentiating the action of drugs handled by the P₄₅₀ system, including benzodiazepines, phenytoin, and morphine (see HISTAMINES AND HISTAMINE ANTAGONISTS).

Drug metabolism also depends on age and sex. The activities of the hepatic biotransforming enzymes are low in the neonatal (and premature) infant. Accordingly, drug toxicity can be unusually apparent, as, for example, chloramphenicol and bilirubin toxicity in newborns, owing to the low activity of conjugation pathways. During various stages of human development, drug metabolism rates can vary considerably. Young children metabolize some drugs, eg, diazoxide and phenobarbital, and eliminate theophylline, faster than adults. With old age, drug metabolism rates generally decline. Hence drug doses for older people should be generally lower than they are for younger ones.

Drug metabolism may also produce toxic materials. Thus, the aromatic hydroxylation of hydrocarbons such as benzpyrene produces the highly reactive and carcinogenic 1,2-epoxides.



Oxidation of acetaminophen yields a reactive quinone intermediate.



Drug Elimination. Drugs are removed from their sites of action through metabolism, storage, and excretion. These processes are not necessarily independent and drugs are frequently metabolized prior to excretion. Indeed, for lipophilic drugs this is virtually a necessity. Drugs are excreted via the kidneys, biliary systems, intestines, and lungs. Renal excretion is the most important. Lactating humans present another route of elimination, ie, via milk, and this can present both advantages and disadvantages to the suckled young. Fecal excretion comprises mainly unabsorbed orally administered drug or biliary-excreted material. Pulmonary excretion deals primarily with volatile anesthetics (see ANESTHETIC AGENTS). Urinary excretion has three components: glomerular filtration, active tubular secretion, and tubular reabsorption. The glomerular membranes in the human kidney filter ca 200 L of fluid per day and excrete ca 1° as urine. Drugs that are not of high molecular weight are filtered in the free state. Thus, drugs tightly bound to proteins are excreted more slowly.

The process of reabsorption depends on the lipophilic–hydrophilic balance of the molecule. Charged and ionized molecules are reabsorbed slowly or not at all. Reabsorption of acidic and basic metabolites is pH-dependent, an important property in detoxification processes in drug poisoning. Both passive and active carrier-mediated mechanisms contribute to tubular drug reabsorption. The process of active tubular secretion handles a number of organic anions and cations, including uric acid, histamine, and choline. Drug metabolites such as glucuronides and organic acids such as penicillin are handled by this process.

Clinical Pharmacokinetics. Clinical pharmacokinetics attempts to define the relationship between drug concentration and therapeutic response. The underlying assumption is that response is proportional to drug concentration at the site of action. This concentration is dependent on many

factors that are frequently pharmacokinetic determinants. The most important factors are defined as clearance, bioavailability, and volume of distribution. Clearance, CL , is defined by

$$DR = CLC_{ss}$$

(1)

where DR represents dosing rate and C_{ss} the steady-state concentration of the drug.

Once the steady-state concentration is known, the rate of drug clearance determines how frequently the drug must be administered. Because most drug elimination systems do not achieve saturation under therapeutic dosing regimens, clearance is independent of plasma concentration of the drug. This first-order elimination of many drugs means that a constant fraction of drug is eliminated per unit time. In the simplest case, clearance can be determined by the dose and the area under the curve (AUC) describing drug concentration as a function of total time:

$$CL_{total} = CL_{renal} + CL_{hepatic} + CL_{other}$$

(2)

This concept becomes important in determining the effects of organ pathology on clearance and on the role of blood flow to individual organs in calculating clearance rate.

The half-life, $t_{1/2}$, for a drug in plasma, ie, the time it takes for the concentration of a drug to be reduced by 50%, is determined by both volume of distribution, V , and clearance:

$$t_{1/2} = 0.693V / CL$$

(3)

The bioavailability of a drug can be defined as the fraction of a dose, F , that reaches the systemic circulation. When $F < 1$,

$$FDR = CLC_{ss}$$

(4)

Pharmacodynamic Aspects of Drug Action

Although the same general principles of chemical specificity apply to all ligand–macromolecular interactions, the term receptor is generally applied to those cellular macromolecules and macromolecular complexes with which ligands, physiological or synthetic, interact both to complex and to initiate a physiological response. Receptors are conveniently viewed as existing in several principal classes, ie, G-protein-coupled receptors, ligand-gated ion channels, voltage-gated ion channels, tyrosine kinase receptors, guanylyl cyclase receptors, and steroid hormone receptors. All of these receptors form homologous classes according to structure and mechanisms of action. G-protein-coupled receptors form a homologous class of membrane proteins characterized by seven transmembrane domains and the ability to couple to guanine (G) nucleotide-binding proteins. Examples of G-coupled receptors are as follows.

Ligand	Receptor
adenosine	A ₁ , A ₂ -adenosine
norepinephrine and epinephrine	1,2-adrenergic
dopamine	D ₁ , D ₂ -dopamine
histamine	H ₂ -histamine
acetylcholine	muscarinic m ₁ –m ₅
5-hydroxytryptamine (5-HT)	5-hydroxytryptamine
prostaglandins	prostaglandin
opiates and enkephalins	opiate
polypeptide hormones	
glucagon	
vasopressin (V ₁ , V ₂)	
thyroid-stimulating hormone (TSH)	
follicle-stimulating hormone (FSH)	
parathyroid hormone (PTH)	
somatostatin (SS)	
neuropeptide Y (NPY)	

Ligand-gated ion channels represent a significant family of ion channels that feature as an integral component of their multimeric subunit organization a receptor site for either acetylcholine (nicotine acetylcholine receptor (AChR)), amino acids including glycine and γ-aminobutyric acid (GABA) (inhibitory transmitters), or glutamic acid (excitatory transmitter). The interaction of the ligand with the endogenous receptor site causes channel opening or closing. Drugs responding to specific ligand-gated ion channels are as follows:

Channel	Drug
nicotinic AChR	curare, gallamine, decamethonium, β-erythroidine, local anesthetics
GABA	benzodiazepines, β-carbolines, avermectin, picrotoxinin
glycine	strychnine
glutamate	phencyclidine, MK 801

Voltage-gated ion channels for the cations Na⁺, K⁺, and Ca²⁺ are similarly multimeric structures. Openings and closings are regulated by changes in membrane potential. Drugs responding to voltage-gated ion channels are as follows:

Channel	Drug
Na ⁺	TTX, DDT, veratridine, scorpion toxins, procaine, lidocaine
K ⁺	quinidine, tolbutamide, diazoxide, glyburide, minoxidil
Ca ²⁺	nifedipine, diltiazem, verapamil, mollusc and spider toxins

An important characteristic of both classes of ion channel is that they possess multiple drug binding sites (Table 2). Many of the channel-active drugs have achieved particular therapeutic importance, including, for example, the Ca²⁺ antagonists, widely used for a number of cardiovascular disorders,

such as hypertension.

Table 2. Drugs Active at Ion Channels

Drug	Property	Channel	Response
nifedipine, verapamil, diltiazem	antagonist	Ca ²⁺	cardiovascular inhibition; hypotensive agents
pinacidil, nicorandil	agonist	K ⁺	cardiovascular and smooth muscle inhibitors
tetrodotoxin	antagonist	Na ⁺	paralysis of nerve and muscle functions
glyburide	antagonist	K ⁺	stimulates insulin release; hypoglycemic agent

The tyrosine kinase receptors serve to recognize a variety of growth factors and growth factor-like agents, including insulin (see INSULIN AND OTHER ANTIDIABETIC AGENTS), platelet-derived growth factor (PDGF), epidermal growth factor (EGF), and colony-macrophage stimulating factor (CMSF). All of these receptors have a common mechanistic link autophosphorylating specific tyrosine residues in the intracellular domain of the receptor itself. A further characteristic of this receptor class is that, upon activation, they mediate both rapid and delayed events. Thus, metabolic responses are typically produced rapidly and within minutes whereas effects on cell growth mediated through deoxyribonucleic acid (DNA) occur according to a time scale of hours to days. Receptors linked to guanylyl cyclase and which catalyze the formation of guanosine triphosphate (GMP) to guanosine-3'5'-cyclic monophosphate (cyclic GMP) include those for atrial natriuretic factor (ANF) and endothelial-derived relaxing factor (EDRF), mediating vasodilatation, and nitric oxide [10102-43-9], NO, or a clearly related derivative.

In marked contrast, the receptors, at which thyroid hormone (see THYROID AND ANTITHYROID PREPARATIONS) and steroids (qv) such as glucocorticoids, sex hormones, and vitamin D exert their long-term effects on cell function through the genetic machinery, are located intracellularly. These agents enter the cell by virtue of their dominantly lipophilic character and interact with a cytosolic receptor. Upon hormone binding, this heteromeric intracellular receptor is activated or converted to a form that binds to specific DNA sequences and alters transcription of a set of specific steroid-responsive genes. Discrete regions upstream of the transcriptional start sequences are known as hormone response elements (HREs). These HREs recognize the hormone–receptor complex and thus mediate transcriptional control. These receptors belong to a family possessing a region rich with carboxy-terminal cysteine and that represents the DNA binding domain. The N-terminal region, which contains the specific hormone binding site, represents the region of greater structural variability.

Structure–Activity Relationships. Until the mid-1980s, the attempted correlation of chemical structure and biological activity was the only available approach to the definition of receptor site structures. The basic assumption in the analysis of structure–activity relationships (SAR) is the existence of a definable mutual complementarity between the structure of the drug and its corresponding binding site. Although this approach has been of considerable value, its application is limited when applied in empirical fashion (22,23). Many drug molecules are flexible structures and, although conformations in the solution and solid states can be determined by spectroscopic and crystallographic methods, these bear no necessary relationship to those adopted at the receptor site. The possibility of mutual conformational adaptation of both the drug and the receptor site during the binding process adds a further complication (24). Furthermore, there may exist multiple drug-binding modes at the receptor such that transitions in binding modes occur at some point in a structurally related series. An additional problem in the quantitative interpretation of SAR is that of the relationship between biological response and drug–receptor interaction. Despite these limitations, SARs have been of great value in providing qualitative concepts of binding site geometry, classifying receptors, furnishing evidence for the existence of new classes of receptor-specific drugs, and generating new and therapeutically effective compounds.

The simplest SARs occur in homologous series of compounds. Thus a linear relationship exists between carbon chain length and biological activity in 1-alkanol-mediated anesthesia (see ANESTHETICS). The activity can be related to the water:cell partition coefficient (25). For other homologous series, however, such linear relationships may not be observed; for example, in the antagonistic activity of α,ω -bistrimethylammonium alkanes at acetylcholine receptors where binding to sites of defined anionic site geometry probably is involved (26).

Relatively unambiguous monotonic SARs also occur where activity depends on the ionization of a particular functional group. A classic example (Fig. 5) is that of the antibacterial sulfonamides where activity is exerted by competitive inhibition of the incorporation of *p*-aminobenzoic acid into folic acid (27). The bell-shaped relationship is consistent with the sulfonamide acting as the anion but permeating into the cell as the neutral species.

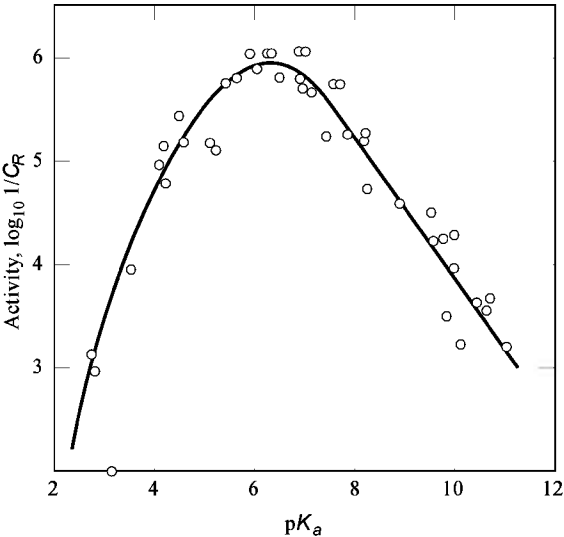


Fig. 5. Relationship between antibacterial activity of sulfonamides (log 1 C_R) and pK_a of sulfonamide NH group where C_R represents the concentration necessary to prevent growth.

Courtesy of American Chemical Society (27).

The SAR is also determined at the level of stereochemistry of interaction (28–30). In principle, three limiting situations can apply to the stereochemistry of drug–receptor interactions: the enantiomers may not differ in activity; the species may differ quantitatively; or they may differ qualitatively. Examples of all three situations are known (28–30). Both enantiomers of dobutamine are inotropic species, but whereas both enantiomers of warfarin are anticoagulant, the activity of (*S*)-warfarin is greater than that of (*R*)-warfarin (see BLOOD, COAGULANTS AND ANTICOAGULANTS). Additionally, (–)-propranolol is a β -blocker and (+)-propranolol is significantly less active. Finally, (–)-sotalol is a β -blocker and (+)-sotalol is a class III antiarrhythmic. The stereoselectivity of drug action is not confined to specific receptor interactions. Stereoselectivity of interaction also occurs in drug transport, protein binding, and metabolism. These stereochemical differences may contribute significantly to the observed stereoselectivity of pharmacologic or therapeutic

response (31,32). The Ca^{2+} channel antagonist verapamil is subject to stereoselective first pass metabolism, whereby the more active (*S*)-verapamil is more rapidly metabolized. The availability of this enantiomer from the clinically administered racemate is greater after intravenous than oral administration. Additionally, verapamil shows modest stereoselective binding, (*S*) > (*R*), to serum albumin (see PHARMACEUTICALS, CHIRAL).

The issue of drug stereoselectivity has become one of both developmental and regulatory significance. In principle, a racemic drug possesses only 50% of the active ingredient, and the rest may have other or interacting pharmacologic activities, which may contribute a metabolic burden or be inert. Over 50% of clinically available drugs have chiral centers and only about 10% of synthetic chiral drugs are marketed in homochiral (enantiomerically pure) form (33). In contrast, drugs that are naturally occurring substances, obtained from or related to naturally occurring molecules, are frequently homochiral.

There is increasing pressure to develop homochiral drugs (34). Growing demands are faced by the pharmaceutical industry in drug development to consider chiral issues in the early preclinical phases of drug design and synthesis.

Often pharmacologic agonist activity decreases and is lost with progressive structural change. A typical example is shown in Figure 6. Increasing *N*-alkyl-substitution in the basic 2-amino-1-(3,4-dihydroxyphenyl)ethanol nucleus, ie, norepinephrine, causes an ultimate loss in activity. However, although inactive as agonists, these higher homologues can interact with the receptor because they act as antagonists of the response induced by the agonists. This antagonism is competitive and, therefore, is consistent with the interaction of these homologues at the receptor, but without the capacity to initiate response. Relatively minor structural changes are frequently sufficient to produce this agonist–antagonist transition.

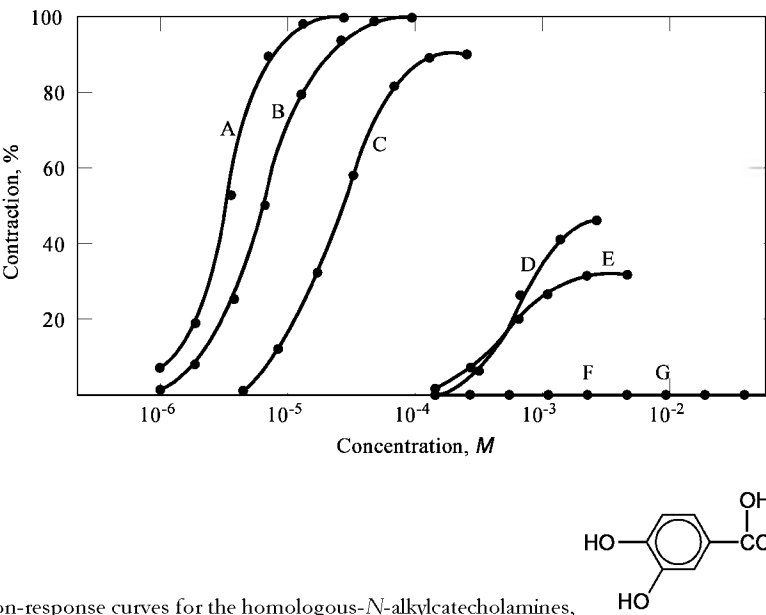


Fig. 6. Cumulative log concentration-response curves for the homologous-*N*-alkylcatecholamines, where A–G correspond to $\text{R} = \text{CH}_3, \text{H}, \text{CH}_2\text{CH}_3, \text{CH}(\text{CH}_3)_2, \text{CH}_2\text{CH}_2\text{CH}_3, \text{C}(\text{CH}_3)_3, \text{and } \text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3$, respectively. There is a gradual change from active to inactive (as agonist) molecules with increasing size of the *N*-alkyl substituent (35).

Increasing attention has been paid to the generation of quantitative structure–activity relationships in which the effects of molecular substitution on pharmacologic activity can be interpreted in terms of the physicochemical properties of the substituents. These approaches are based on the extrathermodynamic analysis of substituent effects (36):

$$\log \frac{k}{k_o} = p\sigma \tag{5}$$

where k_o is the rate or equilibrium constant for an unsubstituted parent compound, k is the same for the substituted compound, σ or σ^- is a parameter describing the electronic effect of the substituent, and p is a proportionality constant. Another substituent constant, ie, the hydrophobic constant, is defined as

$$\log P = \sum \pi; \quad \log \frac{P_o}{P} = \pi \tag{6}$$

where P_o and P are the partition coefficients usually of octanol:water for the unsubstituted and substituted compounds, respectively (36). The use of σ , π , and steric parameters such as E_s has made possible the analysis of multivariate quantitative SARs. A typical example where activity of 51 compounds was determined by ϕ alone is the ability of various organic compounds, eg, alcohols, ethers, and ketones, to produce narcosis in tadpoles:

$$\log \frac{1}{C} = 0.94 \sum \pi + 0.87 \tag{7}$$

where C is the biologically effective concentration (36). The correlation coefficient, r , for equation 7 was 0.97. Because of the structural diversity of the compounds in this case, it is possible that interaction at a specific receptor was not involved. There are also many examples where activity is not linearly dependent on partition coefficients. For the ability of 17 barbiturates to produce hypnosis (36):

$$\log \frac{1}{C} = 0.33\pi^2 + 1.76\pi + 0.93 \tag{8}$$

The correlation coefficient for this equation was 0.994. Such a parabolic dependence of activity on the partition coefficient may reflect partitioning of the drug through several membrane barriers, which enabled the drug to reach its site of action.

Biological activities also may correlate with electronic substituent factors alone, eg, the inhibition of acetylcholinesterase by six diethyl phenyl phosphates (36) gave $r = 0.95$ for

$$\log \frac{1}{C} = 5.77\sigma + 2.71 \tag{9}$$

More commonly, multiparameter correlations can be made. Thus, for the relative sweetness of nine 4-nitro-2-aminobenzenes, $r = 0.97$ (36):

$$\log(\text{relative sweetness}) = 1.03\sigma^+ + 1.43\pi + 1.58$$

(10)

and for the local anesthetic activity of eight 2-diethylaminoethyl benzoates, $r = 0.93$ (36):

$$\log \frac{1}{C} = 0.58\pi - 1.26\sigma + 0.96 \quad (11)$$

The preceding examples of linear free-energy correlations of biological activity with the physicochemical properties of molecules may be used to deduce the types of interaction involved in biological activity and to predict new compounds (37–39). However, generally these predictions are most accurate when they are interpolative rather than extrapolative. Increasing attention is being paid to quantum mechanistic approaches to the definition of SARs. These calculations can provide measurements of nonequilibrium conformational energies, electron densities, and electrostatic potential maps of drug molecules (40).

Advancing technology permits increasing attention to the definition of the three-dimensional structure of the ligand in its bioactive conformation as it binds to the receptor or active site. This bioactive conformation is not necessarily the solution or the crystal structure of the ligand, which is often the most experimentally accessible structure. It is of further critical importance to define the three-dimensional structure of the ligand complexed with its target. This resolution permits not only the understanding of a particular ligand–macromolecule, but also the *in vivo* design of ligand homologues that may have tighter or more selective affinities for the site (40,41).

Considerable effort must be applied to obtaining adequate quantities of the protein target and its structural solution, together with the structural solution of the complexed ligand, either by x-ray or solution nmr techniques. Alternatively, homology modeling may be possible when the structure of a homologue protein is already available. Although many examples of ligand–protein structure determinations are available, some of the most interesting targets, eg, membrane-bound receptors, defy structural solution at the necessary resolution. The examination of the real structure of ligand–receptor complexes should be an increasingly important and integral part of the drug discovery process (Fig. 7) (41).

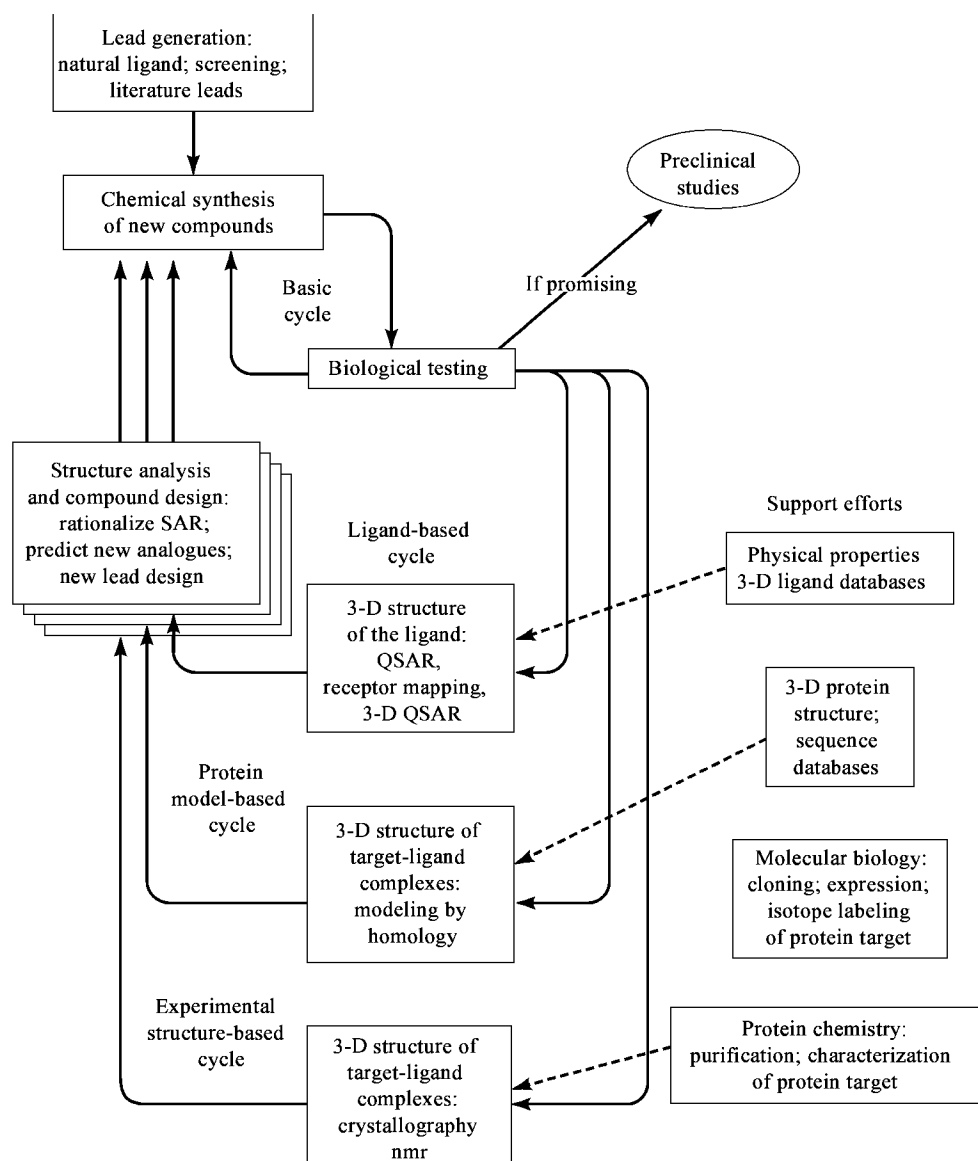


Fig. 7. The cycle of structure-based drug design. The conventional or basic cycle consists of a simple loop between chemical synthesis and biological activity. This cycle is typically initiated through a lead compound. Whereas this cycle remains the core of the drug discovery process, the resources available have been substantially amplified by the indicated developments in structural chemistry.

Courtesy of the American Chemical Society (41).

Quantitative Aspects of Drug–Receptor Interactions. As a general rule, pharmacological responses are graded and a defined relationship exists between the concentration of a drug and the receptor response. This usually is expressed as a concentration–response (A–R) relationship in linear or semilogarithmic coordinates and usually is referred to as a dose–response curve. The shape of these curves (Fig. 8) offers a clear analogy to processes of physical adsorption but, because of the complexity of the sequence of events between drug–receptor interaction and the response, the interpretation of dose–response curves is not simple. A quantitative understanding of drug–receptor interactions is crucial both to the nontrivial interpretation of structure–activity relationships and to the determination of the mechanisms by which drug–receptor complexes initiate pharmacological response.

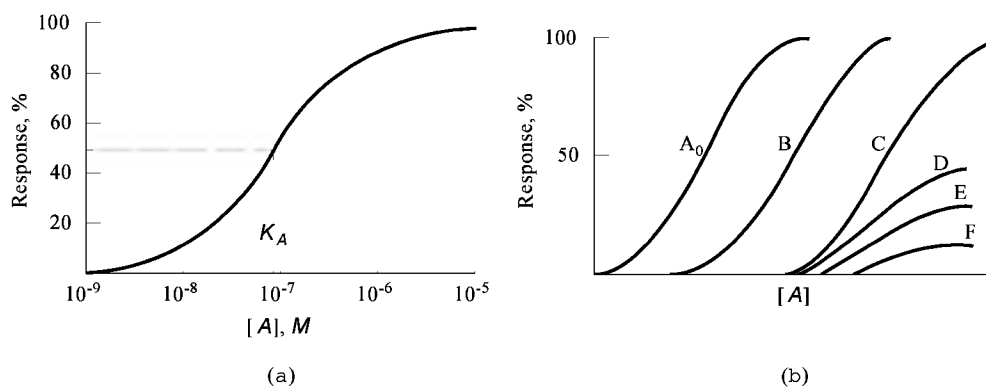


Fig. 8. Agonist A , dose-response curves. (a) For an agonist where a K_A value of $10^{-7} M$ is indicated at the concentration giving 50% response. (b) For an agonist alone, A_0 , and in the presence of increasing amounts of irreversible receptor antagonists, B–F. There is a progressive rightward shift of the dose-response curve prior to reduction of maximum response. This pattern is consistent with the presence of a receptor reserve.

Reaction rates cannot be faster than collision frequencies. Such reactions are diffusion-controlled. The encounter rate, k_e , for drug binding to receptors is

$$k_e = \frac{4\pi N R_{12} D_{12}}{1000} \tag{12}$$

where R_{12} is the sum of the radii of the drug and its binding site, D_{12} is the sum of the self-diffusion coefficients of the two molecular species, and N is Avogadro's number. Such rates frequently approximate $10^6 - 10^8 (M\text{s})^{-1}$ for drug-receptor interactions.

The mechanism by which a drug binds to its specific receptor is important. Effective transmission of chemical information demands accuracy, efficiency, and rapidity. Accuracy can be achieved through the specificity of the molecular architecture of the drug and its receptor. Two extreme situations of drug binding may be visualized and both are consistent with the existence of a structurally and stereochemically defined drug binding site. The drug may interact by presentation of a single favorable binding conformation, ie, lock-and-key principle, or by the zipper model, ie, binding of an initial segment of the drug molecule followed by conformational alignment of the partially bound ligand (42). For drugs that are conformationally mobile, the stringent conformational and orientational demands of the lock-and-key model are likely to reduce substantially the rate of drug-receptor binding. The zipper model may permit a faster rate of drug-receptor interaction because the orientational and conformational requirements for the initial binding of a molecular component are substantially less than for the whole molecule.

In the early twentieth century, the law of mass action was applied to the basic pathway of drug-receptor interaction. Assuming that response, R , is proportional to the concentration of the drug-receptor complex, RA , and that maximum response, R_{max} , occurs when all receptors are occupied (2,6,10), then

$$\frac{R}{R_{\text{max}}} = \frac{[RA]}{[R_{\text{tot}}]} = \frac{[A]}{[A] + K_A} \tag{13}$$

where K_A is the dissociation equilibrium constant for binding of the drug A . Here R_{max} corresponds to complete receptor occupancy. This equation is the equivalent of the Michaelis-Menten expression for enzyme-substrate interactions (1,12). Accordingly, the K_A value of a drug for the receptor is equal to A_{50} , the concentration producing 50% response (Fig. 8a), and a quantitative interpretation of dose-response curves is easily provided. The limiting assumptions made in using equation 13 are (1) binding is reversible; (2) reactants exist only as bound or free species, and degradation, internalization, or other removal mechanisms are not important; and (3) receptor sites are assumed to be of equal affinity and independent in their binding interactions, ie, not cooperative, with ligands.

The original formalism of drug-receptor interactions assumed only two types of drugs, the agonists and the antagonists. However, this was an oversimplification. In many homologous series, such as *N*-alkyltrimethylammonium salts active at cholinergic receptors or *N*-alkylcatecholamines active at adrenergic receptors (Fig. 6), there are clearly ligands that even at saturating concentration produce only partial response (1,10,35). Additionally, the use of irreversibly acting receptor antagonists, eg, 2-halogeneoethylamines such as phenoxybenzamine, which react covalently and eliminate receptors from response generation, did not cause the anticipated reduction of maximum response. Rather, the observed response was frequently a parallel rightward shift of the dose-response curve prior to a depression of response such as that shown in Figure 8b (1,10). These observations indicate that response may not be proportional to receptor occupancy and that spare receptors, ie, receptors in excess of those necessary to generate maximum response, exist. These conclusions have been demonstrated by studies of ligand binding to receptors using radioligands (see RADIOACTIVE TRACERS) and studies of the coupling of receptors to the effector units.

The term intrinsic activity (ia) was defined as a measure of the ability of the drug-receptor complex to generate response. When $ia = 1$, a full agonist is defined; when $ia = 0$, an antagonist is defined. Thus, values $0 < ia < 1$ define partial agonists as follows, where R_A is the response to drug A and R_{max} is the maximum response achieved.

$$\frac{R_A}{R_{\text{max}}} = \frac{[RA]}{[R_{\text{tot}}]} = \frac{[A]}{[A] + K_A} \tag{14}$$

However, maximum receptor occupancy for full agonists was still required for maximum tissue response. The efficacy, e , was introduced through the concept of a stimulus, S , defined as

$$S = \frac{e[RA]}{[R_{\text{tot}}]} = \frac{e[A]}{[A] + K_A} \tag{15}$$

where e is a dimensionless parameter representing the ability of a drug, as the drug-receptor complex, to initiate response (43). Response is some undefined function of stimulus, $f(S)$, which is monotonic and continuous:

$$\frac{R_A}{R_{\text{max}}} = f(S) = f\left(\frac{e[A]}{[A] + K_A}\right) \tag{16}$$

Thus, a drug may produce response either with low efficacy by occupying many receptors or with high efficacy by occupying few receptors. The issues of dealing with agonist–dose response relationships can be complex and reference should be made to detailed texts (44,45).

In contrast, interactions of competitive antagonists and receptors are relatively straightforward. It can be shown by comparing the equal responses provided by an agonist, $[A]$, alone or in the presence of a competitive antagonist, $[A]_B$ so that

$$\frac{[A]_B}{[A]} = 1 + \frac{[B]}{K_B}$$

(17)

The dissociation content for the competitive antagonist, K_B , can be determined without knowing the relationship between receptor occupancy and response. Equation 17 is often written in logarithmic form:

$$\log \frac{[A]_B}{[A]} = 1 + \log [B] - \log K_B$$

(18)

A plot of $\log([A]_B/[A] - 1)$ versus $\log[B]$, called a Schild plot, yields a straight line of unit slope and intercept of K_B , the latter often expressed on a scale analogous to that for pH, so that $pA_2 = \log K_B$ (46–48).

Because K_B values for competitive antagonists represent true dissociation constants, these make possible quantitative interpretations of SARs. Significant use also has been made of K_B values in the quantitative comparison of receptors to determine whether receptors that respond to the same agonists are identical or whether responses produced by different agonists are initiated at the same receptors (44,46). Thus, beta-adrenoceptors in human and guinea pig preparations can be directly compared and selective β_1 and β_2 antagonists quantitated (Table 3).

Table 3. Equilibrium Dissociation Constants for Drug–Receptor Complexes *In Vitro*^a

Antagonists	β-Adrenoceptors			
	Atria		Bronchii	
	Guinea pig	Human	Guinea pig	Human
propranolol	8.5	8.4	8.3	8.6
pindolol	8.7	8.8	8.8	8.6
practolol ^b	6.5	6.4	4.9	4.6
atenolol ^b	7.2	7.0	5.6	5.4
acebutolol ^b	6.5	6.8	5.1	5.1
metoprolol ^b	7.4	7.4	6.1	6.4 ^c

^a Ref. 44.

^b Cardiac selective.

^c Whereas the agreement between the values in humans and guinea pig is close, this is not always so. For example, in human and rodent 5-HT_{1B} receptors, significant pharmacological differences are conferred by a single amino acid residue.

Although the concept of competitive antagonism is well developed in both molecular pharmacology and clinical medicine, many antagonists act noncompetitively. A noncompetitive antagonist acts at a site distinct from the agonist ligand and prevents, by indirect mechanisms, agonist occupancy of its binding site. This concept has been particularly well developed for allosteric proteins where binding or functional behavior of a protein is controlled by equilibrium shifts between active and inactive states according to ligand binding (1,10,44,49).

Direct quantitation of receptor concentrations and drug–receptor interactions is possible by a variety of techniques, including fluorescence, nmr, and radioligand binding. The last is particularly versatile and has been applied both to sophisticated receptor quantitation and to drug screening and discovery protocols (50,51). The use of high specific activity, frequently [³H]- or [¹²⁵I]-labeled, drugs bound to crude or purified cellular materials, to whole cells, or to tissue slices, permits the determination not only of drug–receptor saturation curves, but also of the receptor number, drug affinity, and association and dissociation kinetics either directly or by competition. Complete theoretical and experimental details are available (50,51).

This kind of binding has been demonstrated for many receptor systems and obeys the four anticipated criteria derived from structure–activity relationships: (1) binding should be saturable, which is consistent with cells possessing a finite number of receptors; (2) binding should be reversible, which is consistent with the reversibility of action of most drugs; (3) binding should be observed only in systems known to be sensitive to the drug; and (4) there should exist a correlation between the relative binding affinities and biological activities of drugs that are active in the system (50,51). A good correlation occurs for antagonists between K_D values for biological responses and those from binding studies. A typical example is shown in Figure 9 for atropine-like agents in intestinal smooth muscle, where binding is measured by ligand competition with the specific binding of the potent atropine-like agent, 3-quinuclidinyl benzilate, in tritiated form. Generally, however, a lack of agreement exists when a similar comparison is made for agonist molecules, as is shown in Figure 9b for muscarine-like agonists acting in intestinal smooth muscle (52). Such discrepancies, whereby $K_A/K_D < 1$ and K_A is the pharmacologically determined dissociation constant and K_D the dissociation constant determined from drug binding, are consistent with the existence of complex relationships between agonist concentration–response and saturation curves.

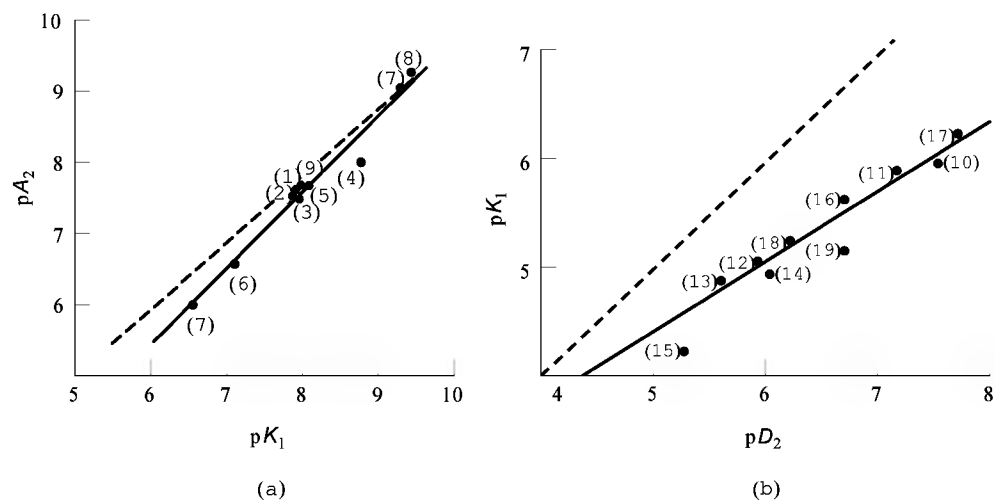


Fig. 9. Correlation between binding and pharmacologic affinities where the dashed lines correspond to the theoretical correlation of 1:1 for a series of muscarinic receptor: (a) antagonists, (1)–(9) and (b) agonists, (10)–(19). Correlation for the antagonists is essentially 1:1, deviating markedly from that relationship for the agonists. The slope is 0.63 at a correlation of 0.96.

Courtesy of the American Society for Pharmacology and Experimental Therapeutics (52). See Table 4.

Table 4. Structures of Muscarinic Receptor Antagonists and Agonists

Structure number	Structure	R	R'	Stereochemistry
<i>Antagonists</i>				
(1) (2) (3) (4) (5) (6)		C H ₅	C H ₅	(R),(S) (R) (S) (2S), (4S)
		C H ₅	C H ₅	(2R), (4R)
		C H ₅	C H ₅	
		C H ₁₁	C H ₅	
		C H ₁₁	C H ₅	
		C ₄ H ₉	C ₄ H ₉	
(7) (8) (9)				(R,S) (R) (S)
<i>Agonists</i>				
(10) (11) (12) (13) (14) (15)		CH ₃ H	H CH ₃	
		Cl ₃ C H H	H CCl ₃	
		CH ₃	H CH ₃	
(16) (17)		H CH ₃		
(18) (19)		H CH ₃		

Nonreceptor-Mediated Drug Action. At least one important class of drugs, the general anesthetics (qv), has been assumed not to owe its therapeutic activities to a specific receptor process. Anesthetic potency shows an excellent linear correlation with partition coefficient and this has been extrapolated to a definition of action at a lipid site. The phospholipids of cell membranes, particularly nerve cells, have been considered as principal targets for general anesthetic action. It has been hypothesized that anesthetics may disrupt phospholipid structure by fluidizing or expanding the cell membrane or by altering the phase relationships of the phospholipids (53,54). However, it is possible that anesthetics bind to hydrophobic sites on proteins and thus affect directly excitable cell behavior (53–55). This latter proposal is consistent both with the activity of the gaseous general anesthetics and with the activity of structurally more complex agents, eg, 3 α -hydroxy-5 α -pregnane-11,20-dione, 3 α -hydroxy-5-pregn-16-ene-11,20-dione, and 1,5-desmethyl-5-cyclohexenylbarbituric acid.

Although most anesthetics are achiral or are administered as racemic mixture, the anesthetic actions are stereoselective. This property can define a specific, rather than a nonspecific, site of action. Stereoselectivity is observed for such barbiturates as thiopental, pentobarbital, and secobarbital. The (S)-enantiomer is modestly more potent (56,57). Additionally, the volatile anesthetic isoflurane also shows stereoselectivity. The (S)-enantiomer is the more active (58). Further evidence that proteins might serve as appropriate targets for general anesthetics come from observations that anesthetics inhibit the activity of the enzyme luciferase. The potencies parallel the anesthetic activities closely (59,60).

It is likely that a principal target of the general anesthetics is neuronal ion channels of both voltage-gated and ligand-gated classes (61,62). Interactions at GABA-mediated inhibitory channels is a significant, but not exclusive, target. Thus, a general anesthetic may have specific but multiple,

rather than nonspecific, sites of action.

Receptor–Effector Coupling. The informational signal initiated by drug–receptor interaction must be translated to biological response. This is activated by a variety of effector-coupling processes that lead to ionic or biochemical changes, including ion channel opening and closing; the formation of second messengers such as cyclic adenosine-3'-5'-monophosphate (cAMP) and inositol-1,4,5-triphosphate (IP₃); and protein phosphorylation through protein kinase A (cAMP-dependent) and protein kinase C (Ca²⁺-dependent), or through autophosphorylation (tyrosine kinase receptors). In these systems, it is increasingly clear that the individual components of a receptor system may be linked in multiple ways. The virtue of this organization lies in the multiple coupling processes permitted beyond a set of components.

These cascades serve as operational amplifiers of the initial ligand–receptor interaction. In each step of the process, amplification by several powers of 10 may occur so that an original signal may be multiplied several millionfold (63).

G-Protein Coupling. The heterotrimeric guanosine triphosphate (GTP) binding proteins, known as G-proteins, are a principal family of proteins serving to couple membrane receptors of the G-protein family to ionic and biochemical processes. This topic is reviewed in References 63–67. The G-proteins are heterotrimers made of three families of subunits, α, β, and γ, which can interact specifically with discrete regions on G-protein-coupled receptors. This includes most receptors for neurotransmitters and polypeptide hormones (see NEUROREGULATORS). G-protein-coupled receptors also embrace the odorant receptor family and the rhodopsin-linked visual cascade.

The underlying coupling mechanisms are defined by the enzymatic activity of the G-protein, that of hydrolyzing GTP, ie, GTPase activity. In the inactive state, the heterotrimeric G-protein is liganded to the diphosphate GDP. Receptor activation reduces the affinity of the α-subunit for GDP and increases the affinity for GTP. The GTP-liganded complex then dissociates to the GTP-bound activated α-subunit and the β- and γ-subunits. These dissociated subunits then interact with the corresponding effectors (Fig. 10). The effectors include adenylyl cyclase, phospholipase C, cGMP phosphodiesterase, some ion channels (K⁺, Ca²⁺), and receptor kinases. These signals may be excitatory or inhibitory according to the class of G-protein, some of which are listed for G-protein-linked adenylyl cyclase:

Stimulation (Gs)	Inhibition (Gi, Go)
β-adrenergic	opiate
H ₂ -histamine	muscarinic
dopamine	α ₁ -adrenergic
polypeptide hormones (glucagon, ACTH, etc)	adenosine (fat cells) A ₁ , prostaglandins (fat cells)
adenosine (platelets, lymphocytes) A ₂	
prostaglandins (platelets)	polypeptide hormones
serotonin (5-HT _{1α})	somatostatin, neuropeptide Y, atriopeptin

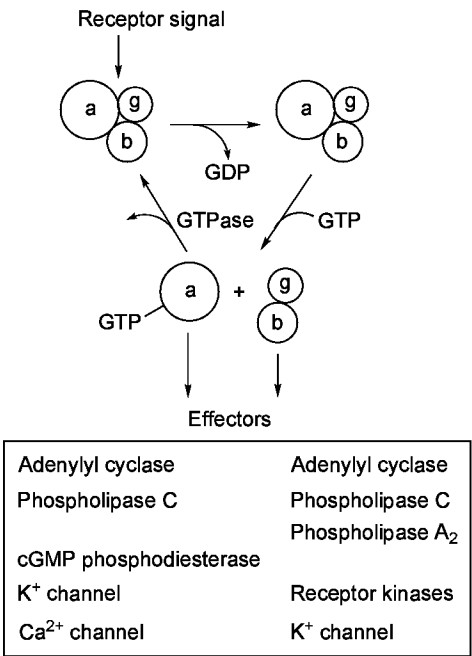


Fig. 10. The receptor–G-protein sequence. An activated receptor interacts with the trimeric GDP-ligated receptor to cause an interchange of GDP by GTP and dissociation into the activated Gα–GTP (left) and Gβγ (right) subunits. These then interact with a variety of effectors. The purpose of the activated receptor is to act as a switch for the G-protein complex.

A critical component of the G-protein effector cascade is the hydrolysis of GTP by the activated α-subunit (GTPase). This provides not only a component of the amplification process of the G-protein cascade (63) but also serves to provide further measures of drug efficacy. Additionally, the scheme of Figure 10 indicates that the coupling process also depends on the stoichiometry of receptors and G-proteins. A reduction in receptor number should diminish the efficacy of coupling and thus reduce drug efficacy. This is seen in Figure 11, which indicates that the ability of the muscarinic drug carbachol [51-83-2] to inhibit cAMP formation and to stimulate inositol triphosphate, IP₃, formation yields different dose–response curves, and that after receptor removal by irreversible alkylation, carbachol becomes a partial agonist (68).

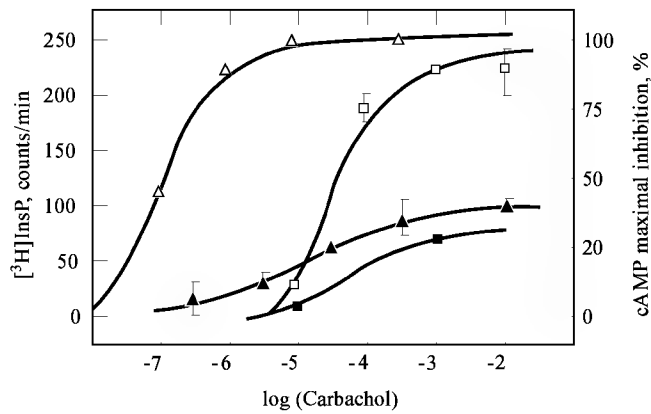


Fig. 11. Dose-response curves for (Δ,▲) inhibition of cyclic AMP formation and stimulation of IP₃ formation by carbachol (Δ,□) before and (▲,■) after reduction of receptor number by irreversible alkylation; (carbachol) is in M. Error bars () are shown for some studies.

Courtesy of the American Society of Pharmacology and Experimental Therapeutics (68).

The ability of receptors to couple to G-proteins and initiate GTPase activity may also be independent of ligand. Thus, specific mutations in α- and β-adrenergic receptors have led to receptors that mediate agonist-independent activation of adenylyl cyclase (69,70). These mutations presumably mimic the conformational state of the ligand-activated receptor when they are activated conventionally by ligands.

GTPase activity is also associated with the Ras protein family. These small proteins act as binding switches, turning GTPase on and off, and are also regulated by other proteins, including GTPase activating proteins (GAP), guanine-nucleotide exchange factors (GEF), as well as guanine-nucleotide dissociation inhibitors (GDI) and stimulators (GDS) (71). The Ras gene family is an extended family of proteins involved in cell growth and behavior. In mammalian cells, mutational activation of the Ras proteins leads to oncogene function, and constitutive activation of this pathway leads to malignant transformation (72).

The principal intracellular messengers derived from activation of G-protein-coupled receptors are cAMP and IP₃. cAMP may be degraded by phosphodiesterase (PDE) or it may activate cAMP-dependent protein kinase (PKA). The activation of this enzyme involves dissociation of the inactive form (R₂C₂) into the active form which subsequently phosphorylates specific proteins (Fig. 12). In contrast, IP₃, one of the products of receptor-mediated phospholipase C breakdown of phosphatidylinositol (PI) (Fig. 13), acts on specific receptors in the endoplasmic reticulum to release Ca²⁺ from intracellular sources. The other product of PI turnover is a 1,2-diacylglycerol that activates protein kinase C (PKC). This is also the receptor for the tumor-promoting phorbol esters (73). These diacylglycerols can be cleaved by monoacyl- or diacylglycerol kinases to yield arachidonic acid, a precursor to the prostaglandins (qv) and thromboxanes.

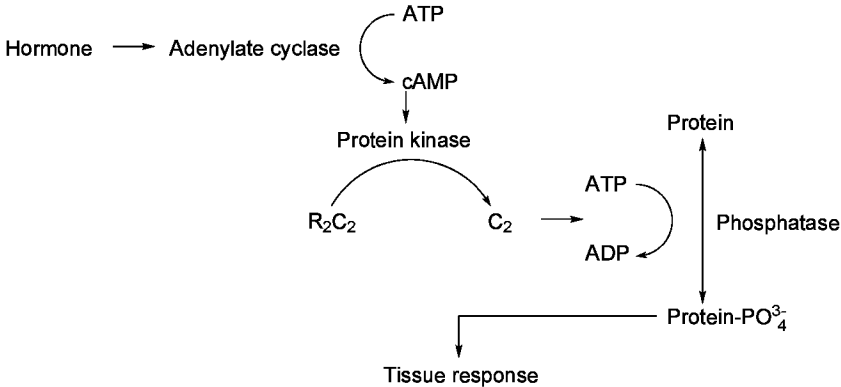


Fig. 12. The hormone-activated adenylyl cyclase cascade.

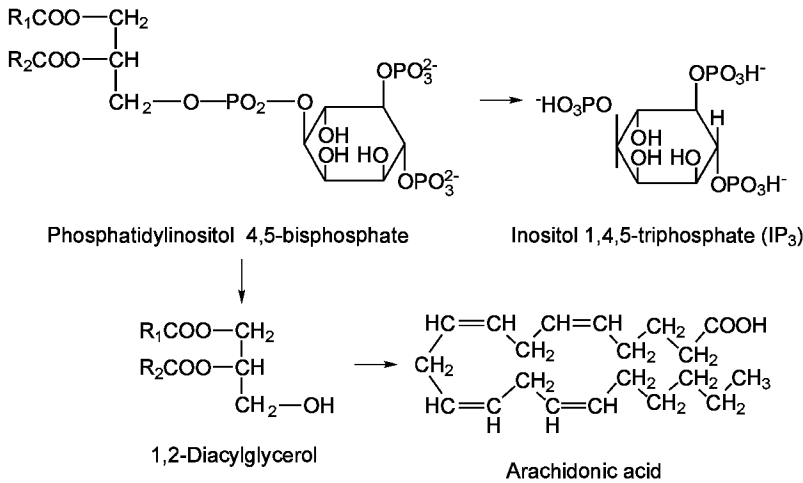


Fig. 13. The phosphatidylinositol pathway.

Ion Channels. The excitable cell maintains an asymmetric distribution across both the plasma membrane, defining the extracellular and intracellular environments, as well as the intracellular membranes which define the cellular organelles. This maintained asymmetric distribution of ions serves two principal objectives. It contributes to the generation and maintenance of a potential gradient and the subsequent generation of electrical currents following appropriate stimulation. Moreover, it permits the ions themselves to serve as cellular messengers to link membrane excitation and cellular

response (74). In some instances, the current itself may be the response, as, for example, in the electric organ of electric fishes. In most instances, however, the current serves to initiate or modulate another cellular response, including propagation of impulses in nerve fibers, and alteration of the sensitivity of membranes to other stimuli or coupling to cellular responses such as contraction and secretion. In the latter examples, a role for calcium is particularly prominent because Ca^{2+} can serve as both a current-carrying and a messenger species (75).

Regulation of ion channels by drugs may have excitatory or inhibitory effects according to the channels affected (Fig. 14). Thus, activation of sodium or Ca^{2+} channels represents stimulatory events, driving the membrane potential to the depolarized equilibrium potential for these ions. In contrast, activation of K^{+} or Cl^{-} channels generally increases membrane potential representing an inhibitory signal. Conversely, antagonists at Na^{+} and Ca^{2+} channels generally have inhibitory effects, whereas antagonists at K^{+} and Cl^{-} channels are excitatory (Table 5).

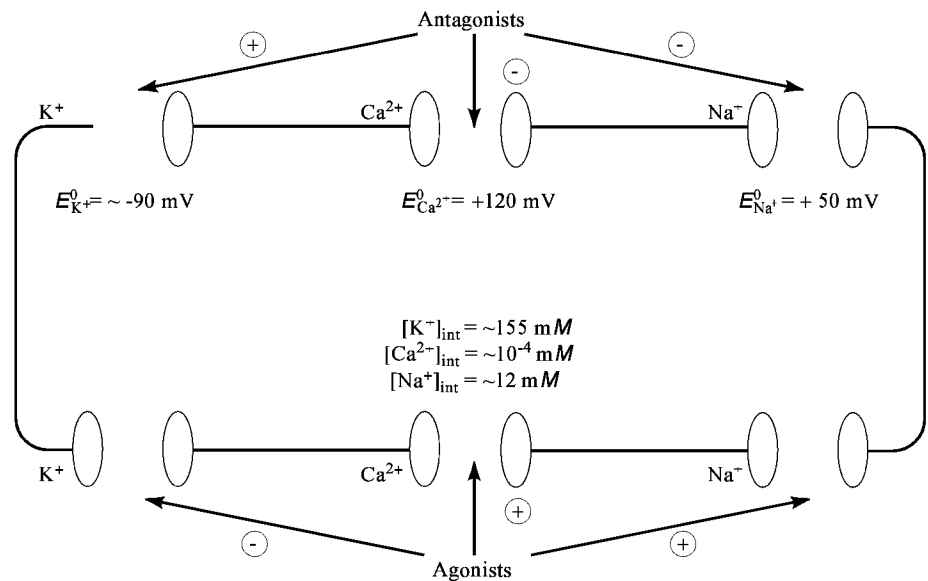


Fig. 14. The cellular ionic environment depicting representative intracellular ionic concentrations and the equilibrium potentials, E^0 , for individual ions. Excitatory and inhibitory events are represented by $-$ and $+$, respectively. Thus, K^{+} channel agonists and antagonists are inhibitory and excitatory, respectively; Ca^{2+} channel antagonists and activators are inhibitory and excitatory, respectively.

Table 5. Drug Activity at Ion Channels

Ion channel	Drug activity	
	Inhibitory	Stimulatory
calcium	verapamil; nifedipine; diltiazem	Bay K 8644
potassium	glibenclamide; quaternary ammonium salts; charybdotoxin and other toxins	minoxidil; pinacidil
sodium	tetrodotoxin; saxitoxin; local anesthetics; DDT	veratridine

Channels may be regulated exclusively by electrical or chemical signals corresponding to purely voltage-gated or ligand-gated channels, respectively. However, receptor activation may alter the operation of a potential-dependent channel, for example, by phosphorylation processes. Many ligand-sensitive ion channel processes show limited dependence on membrane potential because of the influence of electric field on the dipoles and orientations of membrane proteins or on the ligands themselves. Receptor processes may alter ion channel activity directly or indirectly. Thus, the nicotinic acetylcholine receptor consists of five subunits, two of which comprise the acetylcholine binding sites and all of which comprise the ion channel (75). This constitutes a direct linkage. In many instances, however, the link is indirect via a second messenger, such as cAMP or cGMP, derived from stimulation of adenylate and guanylate cyclases or inositol polyphosphate derived from phospholipase C-stimulated hydrolysis of inositol phospholipids (76). Regardless of regulatory mechanism, ion channels may be regarded as allosteric enzymes. The function is to accelerate the transit of ions across an essentially impermeable barrier and to be responsive to a variety of heterotropic signals (74).

Ion channels may be regarded as pharmacological receptors frequently possessing a multiplicity of drug binding sites (77,78). These sites may be for endogenous physiological regulators or for endogenous or synthetic agents. This is illustrated in terms of the ligand-gated γ -aminobutyric acid channel (Fig. 15) and the voltage-gated Ca^{2+} channel (Fig. 16). Ion channels may be regarded, in a limiting sense, as probabilistic devices existing in two states, open (conducting) and closed (nonconducting). The current flowing is given by

$$I = N_f \cdot P_o \cdot i \tag{19}$$

where N_f is the number of functional channels, P_o is the opening probability, and i is the unitary, ie, single channel, current. Further,

$$N_f = N_{\text{tot}} \cdot P_f \tag{20}$$

where N_{tot} is the total number of ion channels and P_f is the probability that a channel is available. Drugs may affect the probability of channel availability or opening either by biochemical events, including G-protein interaction and phosphorylation, or biophysically by altering the voltage dependence of activation or inactivation.

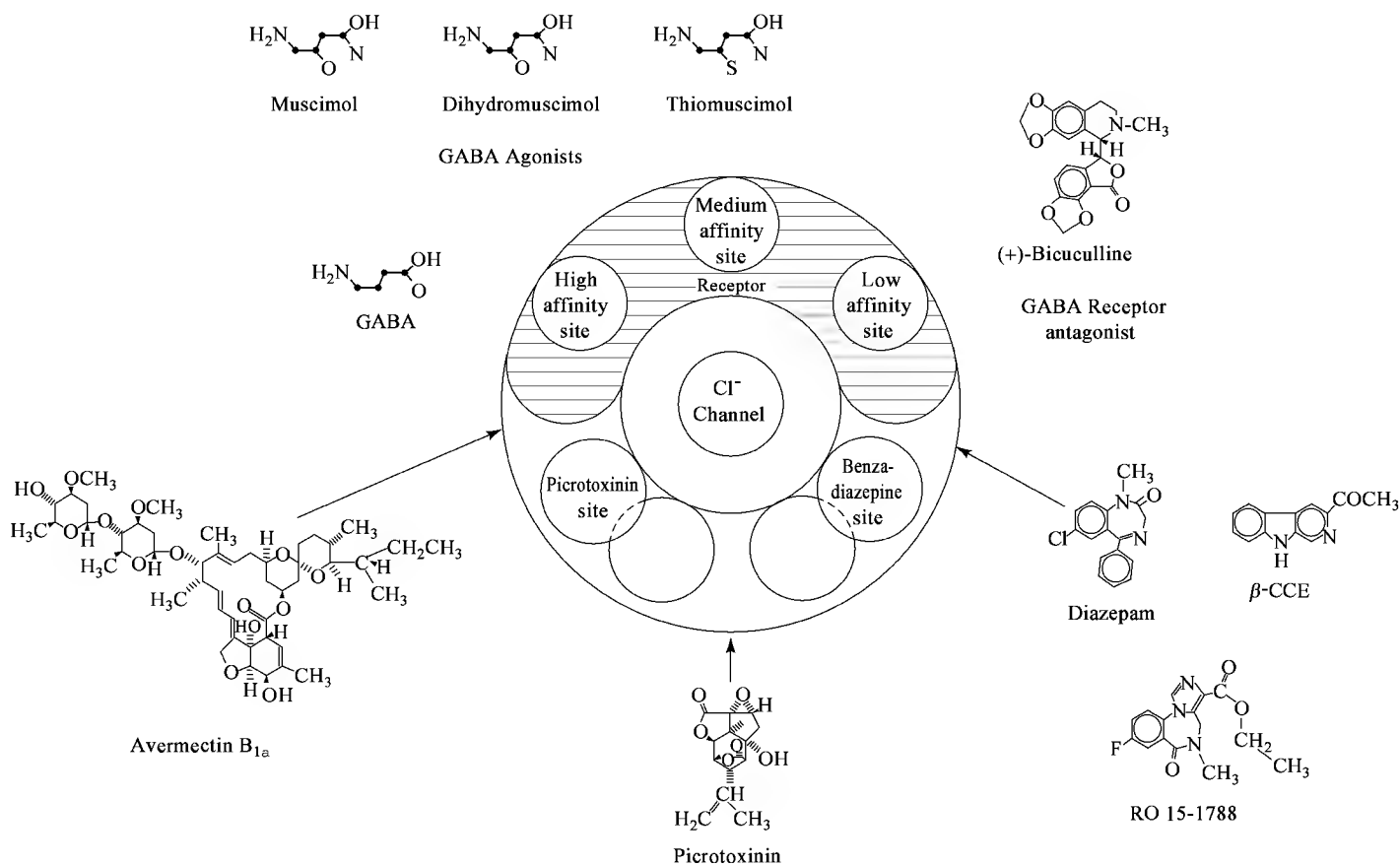


Fig. 15. Drug binding sites associated with the GABA_A receptor–channel complex where (—) represents the carbon backbone of GABA agonists.

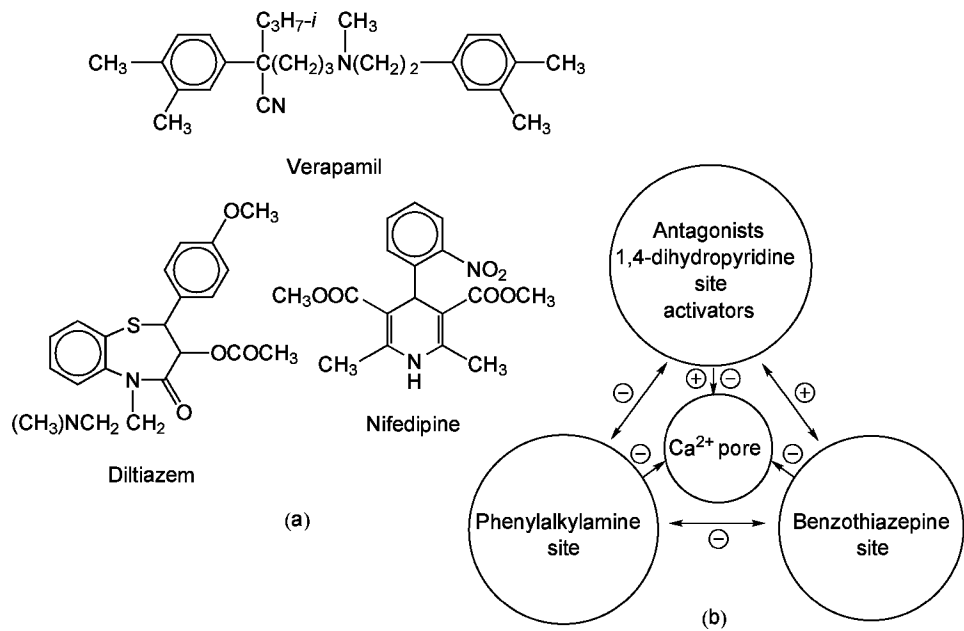


Fig. 16. (a) Structural formulas of the first-generation Ca²⁺ channel antagonists indicating chemical heterogeneity consistent with interaction at discrete drug binding sites associated with (b) the voltage-gated L-type of Ca²⁺ channel. The second-generation 1,4-dihydropyridine antagonists (amlodipine, felodipine, isradipine, nicardipine, and nimodipine) interact at the nifedipine site.

Tyrosine Kinase Receptors. The polypeptide growth factors control cell proliferation, differentiation, and survival (79). Several distinct subfamilies of receptor tyrosine kinases exist and at least nine have been characterized (80). These include families for epidermal growth factor, insulin and insulin-related factors, fibroblast growth factors, and neurotrophin receptors such as nerve growth factor and brain-derived neurotrophic factor. All of these receptors have kinetics that share certain fundamental signaling properties. Ligand binding to the extracellular domain activates a tyrosine kinase of the cytoplasmic domain. Subsequently, a variety of downstream signaling molecules are activated. These include phospholipase C, GTPase activating factor (GAP), Ras, and MAP kinases (79,81,82).

Guanylyl Cyclase Receptors. Cyclic GMP concentrations (cGMP) rise in response to a number of cell signals (83). Membrane-associated guanylyl cyclase catalyzes the conversion of guanosine triphosphate (GTP) to cGMP (84). This enzyme resembles in organization the tyrosine kinases having an intracellular protein kinase-like domain and a cyclase catalytic domain. The enzymes are activated by several distinct species that include atrial natriuretic peptide (ANF) and peptides related to the heart-stable enterotoxins.

In contrast, the soluble guanylyl cyclases are regulated by nitric oxide and NO-forming drugs through the Ca²⁺ calmodulin-dependent nitric oxide synthase (85,86). The soluble cGMP, derived either from the particulate or soluble forms of the enzyme, now functions as a second messenger to interact with a number of discrete pathways, including cGMP-gated ion channels, cGMP-inhibited cAMP phosphodiesterase, cGMP-stimulated cAMP phosphodiesterase, and cGMP-dependent protein kinases.

Receptor Regulation and Defects. Specific recognition and the initiation of response are the accepted attributes of the drug–receptor interaction. However, target cells can alter on both short- and long-term time scales their sensitivity to drugs. Such regulation, achieved by altering the

number and or affinity of receptors, is well established for all receptor systems and can be viewed as an integral component of the drug–receptor interaction. In this view, subsequent to the formation of the drug–receptor complex with agonist, the continued existence of the drug–receptor complex may lead to one or more phases of desensitization, according to which there may occur initially transient and subsequently prolonged phases of reduced or lost sensitivity (1,86). Occupancy by antagonist, in contrast, leads to an increased number of receptors and increased drug sensitivity. This phenomenon may contribute to clinical rebound during abrupt withdrawal from drugs, including β -blockers (87–89). Additional to this homologous regulation, receptor sensitivity may be controlled through heterologous influences, whereby hormones, including thyroid and corticosteroids, regulate other receptors. These regulatory events are made possible because pharmacologic receptors, in common with other cellular components, are in dynamic balance between synthesis and degradation. This balance is sensitive to a number of influences that include agonist and antagonist presence.

There are probably several processes that contribute to the total desensitization process and these may be directed homologously (to own receptor) or heterologously (to other receptor). Additionally, the influences may be directed at the receptor itself and affect only that receptor, ie, specific desensitization, or may affect other receptor processes as well, ie, nonspecific desensitization.

A number of distinct processes underlie the several receptor regulatory events and these may be distinguished in part by the time scale on which they occur. Cells frequently exhibit several desensitization events. Rapid desensitization processes frequently occur with ion channels of both the ligand-gated and voltage-gated families. On this time scale, channels may open and subsequently close in the maintained presence of the drug or stimulus in seconds or less (74). Such a process is usually rapidly reversible and involves the formation of a closed channel state. For G-protein coupled receptors, and particularly for the β -adrenergic receptor, the desensitization process has been shown to involve several stages. Both protein kinase A (PKA), activated through cAMP, and β -adrenergic receptor kinase (bARK) are involved, and the receptor phosphorylation, at different sites, uncouples the receptor and G-protein (90,91). At low agonist concentrations, phosphorylation is principally through the PKA pathway; at high agonist concentrations, both the PKA and bARK pathways are involved. Continued occupancy by agonist leads to a second phase of desensitization in which the receptor is sequestered or transiently and reversibly internalized in vesicular form (Fig. 17). Further occupancy leads to down-regulation proper in which the receptors are internalized and reprocessed through the lysosomal machinery. Similar events, but differing in detail, seem likely to occur for some other receptors.

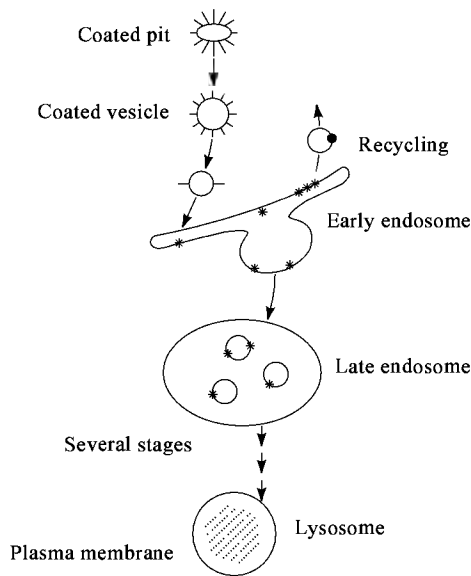


Fig. 17. The receptor-internalization process via a coated pit in which receptors are internalized and processed in a multistep pathway before being recycled or reprocessed.

Specific processes of endocytosis occur that serve to translocate receptor, ligand, or both into the cell interior and which function also as part of the cell's physiological control mechanisms (92–95). This endocytotic process described for low density lipoprotein (LDL) entry into the cell via the LDL receptor (Fig. 18) is used by a large number of receptors and ligands, including LDL, asialoglycoproteins, transferrin, class I and II major histocompatibility complex (MHC) molecules, epidermal growth factors (EGF), and immunoglobulin G (IgG). In some instances, the receptor recycles and the ligand (LDL, viruses, peptide hormones, asialoglycoproteins) is degraded. In others, both receptor and ligand (transferrin) recycle or both are degraded (EGF).

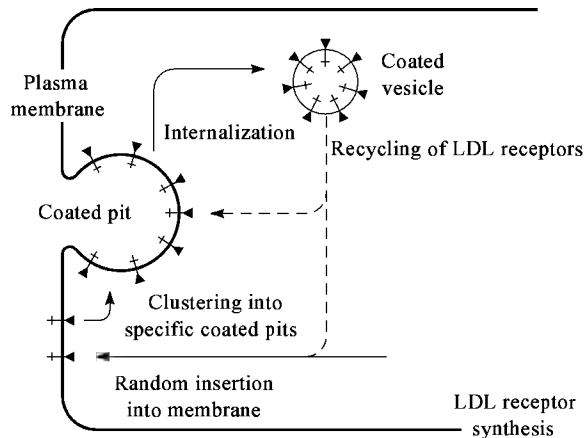


Fig. 18. Schematic representation of cycling of low density lipoprotein (LDL) receptors from the plasma membrane to the cell interior.

Courtesy of Annual Reviews Inc. (93).

An increasing number of diseases are known to be linked to defects in receptor structure, function, or coupling. The defects may lie at several locations: in the structure of the receptor, which may alter its ability either to bind drugs, to be inserted into the membrane, or to couple to effectors (including G-proteins); in the coupling protein; or in the presence of autoantibodies, which can proceed to activate, block, or lyse the receptors and its components (96–99).

Autoantibodies are directed against nicotinic acetylcholine receptors in myasthenia gravis, resulting in receptor loss, skeletal muscle paralysis, and dysfunction (100). In addition, antibodies directed against voltage-gated Ca^{2+} channels produce similar neuromuscular dysfunction of Lambert-Eaton

syndrome (101), and antibodies in amyotrophic lateral sclerosis (Lou Gehrig’s disease) may contribute to muscle wastage and loss by overstimulating the Ca^{2+} channels (102).

In type A behavior in humans, an association has been suggested to correlate with the ratio of peripheral α_2 - and β_2 -adrenoceptor densities (103). Catecholamine receptors are well established to be altered by a variety of homologous and heterologous influences (104). Thus, in hyperthyroidism, there is an increased level of sympathetic activity associated with increased expression of α - and β -adrenoceptors.

Structural defects at the receptor level are determinant for a number of receptor diseases. In nephrogenic diabetes insipidus, where patients void large volumes of dilute urine even in the presence of vasopressin (antidiuretic hormone) (105), the disease is linked to mutations in three discrete regions of the G-protein-linked vasopressin (V_2) receptor (106,107).

Pseudohypoparathyroidism is characterized by end-organ resistance to parathyroid hormone (98,108). This disease takes various forms, including Albright’s hereditary osteodystrophy, which has unusual physical features and a generalized resistance to G-protein-linked hormones that function through cAMP as a second messenger. This defect is associated with a deficiency in the levels of the α -subunit of G_i (109). Because this defect may be generalized, such patients also have olfactory dysfunction (110).

Defects in the LDL receptor have been particularly well explored as a basis of the disease familial hypercholesterolemia (93,111). A number of defects that collectively impair LDL receptor trafficking, binding, or delivery underlie this disease where LDL and serum cholesterol rise to levels that mediate early cardiovascular mortality. Studies of the population distribution of this defect can determine the source of the original mutation. Thus, in Quebec, about 60% of the individuals suffering from familial hypercholesterolemia have a particular 10-kilobase deletion mutation in the LDL gene (112). This may have arisen from an original founder of the French Canadian settlement in the seventeenth century.

Cystic fibrosis, a disease of the Caucasian population, is associated with defective Cl^- regulation and is essentially a disorder of epithelial cells (113,114). The defect arises at several levels in the Cl^- ion transporter, ie, the cystic fibrosis transmembrane regulation (CFTR), and is associated with defective Cl^- transport and defective processing, whereby the protein is not correctly incorporated into the cell membrane. The most common mutation, affecting approximately 60% of patients, is termed F 608 and designates the loss of phenylalanine at this position. This mutation appears to be at least 50,000 years old, which suggests that its survival may have had evolutionary significance (115).

Components of Drug Action and Responses to Drugs. The response to a drug can vary among race, gender, and age groups. It may vary according to disease state and age, and it may vary according to the time of administration (1,2). These factors may have several origins, including (1) compliance, the ability or desire of the subject to take a drug according to a specific regimen; (2) pharmacokinetic, disease-, age-, race-, and gender-based factors that contribute to variable absorption, distribution, metabolism, and excretion of a drug; and (3) pharmacodynamic, disease-, age-, race-, and gender-based factors that contribute to variable drug–receptor interactions.

These same responses may also be classified as follows.

- (1) Geriatric factors: a variety of factors, both pharmacokinetic and pharmacodynamic, that contribute to variable drug responses in the elderly. These responses are not seen for every class of drug. Thus, the depressant effects of the glycosides also appear to increase with aging (116,117).
 - (2) Pharmacogenetics: the responses to drugs may be significantly different according to heritable factors that can modulate pharmacodynamic or pharmacogenetic factors (118). Atypical cholinesterase occurs in about 1 in 2000 Caucasians and is associated with a markedly reduced sensitivity to hydrolysis of the muscle-relaxant cholinesterase. Similarly, the reduced sensitivity to the anticoagulant warfarin is associated with a reduced receptor affinity.
 - (3) Racial differences: differences in racial sensitivity to drug action are quite common, although the origins of most of these differences have not been determined. In Chinese individuals, the plasma levels of α_1 -glycoprotein are lower and drug binding may be lower (119). Several differences have been noted in racial sensitivity to antihypertensive drugs. Males of Chinese descent are more sensitive to the β -antagonist propranolol than are whites (120). This difference arises in part from decreased plasma protein binding of propranolol in Chinese individuals. Similarly, in a comparison of the efficacies of a group of antihypertensive drugs, significant race-related differences were observed. Whites are more sensitive to ACE inhibitors; blacks more sensitive to Ca^{2+} blockers (121).
 - (4) Hormonal factors: drug action can be significantly altered according to hormonal status. Thus, serum levels of α_1 -acid glycoprotein vary according to the phase of the human menstrual cycle and set the stage for differences in drug binding (122). During pregnancy, there are significant reductions in the levels of plasma proteins and thus the binding of both basic and acidic drugs is reduced (123).
 - (5) Time factors: increasingly, it is realized that drug effects vary according to biological timing and endogenous periodicities (124). Many examples are known; eg, the effects of ranitidine, a histamine H_2 receptor antagonist, on gastric pH were greatest at night, and the anticancer drugs 5-fluorouracil and adriamycin achieved different plasma levels over a period of time.
- Factors such as age, gender, race, sex, disease, and time are all integral to considerations of drug efficacy and administration. These are expected to become routine determinants of the analysis of clinical drug action. The study of drug action has moved from phenomenological descriptions at the beginning of the 1900s to quantitative ones. The increasing volume of both chemical (structural) and biological (genetic) information should lead to refinements of the understanding of drug action.

BIBLIOGRAPHY

"Pharmacodynamics" in *ECT* 3rd ed., Vol. 17, pp. 346–372, by D. J. Triggle, State University of New York at Buffalo.

1. W. B. Pratt and P. Taylor, eds., *Principles of Drug Action: The Basis of Pharmacology*, 3rd ed., Churchill Livingstone, New York, 1990.
2. A. G. Gilman and co-workers, eds., *The Pharmacological Basis of Therapeutics*, 8th ed., Pergamon Press, New York, 1990.
3. M. D. Houslay and K. K. Staley, *Dynamics of Biological Membranes*, John Wiley & Sons, Inc., New York, 1982.
4. M. Mathieu, *New Drug Development: A Regulatory Overview*, Parexel, Waltham, Mass., 1994.
5. J. Parascondale, *Trends Pharmacol. Sci.* **11**, 189 (1980).
6. A. J. Clark, *The Mode of Action of Drugs on Cells*, Arnold, London, 1937.
7. P. Ehrlich, *The Collected Papers of Paul Ehrlich*, Vol. 2, Pergamon Press, Oxford, U.K., 1957, p. 178.
8. J. N. Langley, *J. Physiol.* **1**, 367 (1878).
9. J. N. Langley, *Proc. Roy. Soc. Lond.* **B78**, 170 (1906).
10. D. J. Triggle and C. R. Triggle, *Chemical Pharmacology of the Synapse*, Academic Press, New York, 1977, Chaps. 2–3.
11. M. F. Greaves, *Cellular Recognition*, Chapman and Hall, London, 1975.
12. A. Albert, *Selective Toxicity*, 7th ed., Chapman and Hall, London, 1981.
13. J. S. Morley, *Ann. Rev. Pharmacol. Toxicol.* **20**, 81 (1980).
14. E. Costa and A. Guidotti, *Life Sciences* **49**, 325–344 (1991).
15. L. A. Matsuda and co-workers, *Nature*, **346**, 561–564 (1990).
16. W. A. Devane and co-workers, *Science*, **258**, 1946–1949 (1992).
17. D. M. Evans and co-workers, *J. Pharmacol. Exp. Ther.* **268**, 1271–1277 (1994).
18. L. Shargel and A. B. C. Yu, *Applied Biopharmaceutics and Pharmacokinetics*, 2nd ed., Appleton-Century Crofts, Norwalk, Conn., 1985.
19. P. A. Routledge, *Br. J. Clin. Pharmacol.* **22**, 499–506 (1986).
20. R. D. Blakely, *Curr. Opin. Psychol.* **5**, 69–73 (1992).
21. S. Cholerton, A. K. Daly, and J. R. Idle, *Trends Pharmacol. Sci.* **13**, 434 (1992).

22. Y. Martin, *Methods Enzymol.* **203**, 587 (1991).

23. H. Kubinyi, *QSAR: Hansch Analysis and Related Approaches*, VCH Weinheim, New York, 1993.

24. R. A. Wiley and D. H. Rich, *Med. Res. Revs.* **13**, 327 (1993).

25. A. Meyer and H. Hemmi, *Biochim. Zeit.* **277**, 39 (1935).

26. W. D. M. Paton and E. Zaimis, *Brit. J. Pharmacol.* **4**, 381 (1949).

27. P. H. Bell and R. O. Roblin, *J. Amer. Chem. Soc.* **64**, 2905 (1942).

28. A. F. Casy, *The Steric Factor in Medicinal Chemistry*, Plenum Press, New York, 1993.

29. R. Crossley, *Tetrahedron*, **48**, 8155–8178 (1992).

30. R. R. Ruffolo, *Tetrahedron*, **47**, 9953–9980 (1991).

31. F. Jamali, R. Mehvar, and F. M. Pasutto, *J. Pharm. Sci.* **78**, 695–715 (1989).

32. T. A. G. Noctar, "Enantioselective Binding of Drugs To Plasma Proteins", in I. W. Wainer, ed., *Drug Stereochemistry*, 2nd ed., Marcel Dekker, Inc., New York, 1993.

33. S. C. Stinson, *Chem. Eng. News*, **27**, 38 (1993).

34. Chirality, **4**, 338–340 (1992).

35. E. J. Ariens, *Molecular Pharmacology*, Vol. 1, Academic Press, Inc., New York, 1964.

36. C. Hansch and A. J. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, John Wiley & Sons, Inc., New York, 1979.

37. P. Mager, *Multidimensional Pharmacology*, Academic Press, Inc., New York, 1984.

38. C. A. Ramsden, *Quantitative Drug Design*, Pergamon Press, Oxford, U.K., 1990.

39. C. Silipo and A. Vittoria, *QSAR: Rational Approaches to the Design of Bioactive Compounds*, Elsevier, Amsterdam, the Netherlands, 1991.

40. T. J. Perun and C. L. Propst, *Computer-aided Drug Design*, Marcel Dekker, Inc., New York, 1989.

41. J. Greer and co-workers, *J. Med. Chem.* **37**, 1035–1054 (1994).

42. A. S. V. Burgen, G. C. K. Roberts, and J. Feeney, *Nature*, **253**, 753 (1975).

43. R. P. Stephenson, *Brit. J. Pharmacol.* **11**, 379 (1956).

44. T. Kenakin, *Pharmacological Analysis of Drug Receptor Interaction*, 2nd ed., Raven, New York, 1993; D. Oksenberg and co-workers, *Nature*, **360**, 161 (1992).

45. J. W. Black and P. Leff, *Proc. Roy. Soc. Lond. Ser. B*, **220**, 141 (1983).

46. O. Arunlakshana and H. O. Schild, *Brit. J. Pharmacol.* **14**, 48 (1959).

47. R. B. Badow, *Quantitative Aspects of Chemical Pharmacology*, Groom Helm, London, 1980.

48. R. J. Tallarida and L. S. Jacob, *The Dose–Response Curve in Pharmacology*, Springer Verlag, New York, 1979.

49. J. Monod, J. Wyman, and J. P. Changeux, *J. Mol. Biol.* **12**, 88 (1965).

50. H. I. Yamamura, S. J. Enna, and M. J. Kuhar, eds., *Methods in Neurotransmitter Receptor Analysis*, Raven, New York, 1990.

51. E. C. Hulme, ed., *Receptor-Ligand Interactions*, IRL Press, Oxford, U.K., 1992.

52. R. S. Aronstam, M. E. Eldefrawi, and D. J. Triggle, *Mol. Pharmacol.* **15**, 227 (1979).

53. N. P. Franks and W. R. Lieb, *Nature*, **274**, 339 (1978).

54. C. D. Richards and co-workers, *Nature*, **276**, 775 (1978).

55. M. J. Halsey, R. A. Miller, and J. A. Sutton, *Molecular Mechanisms in General Anesthesia*, Churchill Livingstone, Edinburgh, Scotland, 1974.

56. P. R. Andrews and L. C. Mark, *Anesthesiology*, **57**, 314–320 (1982).

57. J. A. Richter and J. R. Holtman, *Prog. Neurobiol.* **18**, 275–319 (1982).

58. B. Harris, E. Moody, and P. Skolnick, *Eur. J. Pharmacol.* **217**, 215–216 (1992).

59. N. P. Franks and W. R. Lieb, *Nature*, **310**, 599–601 (1984).

60. G. W. J. Moss, N. P. Franks, and W. R. Lieb, *Proc. Nat. Acad. Sci.* **88**, 134–138 (1991).

61. N. P. Franks and W. R. Lieb, *Trends Pharmacol. Sci.* **8**, 169–174 (1987).

62. N. P. Franks and W. R. Lieb, *Nature*, **367**, 607–614 (1994).

63. T. D. Lamb and E. N. Pugh, *Trends Neurosci.* **15**, 291–298 (1992).

64. A. G. Gilman, *Ann. Rev. Biochem.* **56**, 615–649 (1987).

65. H. R. Bourne, D. A. Sanders, and F. McCormick, *Nature*, **349**, 117–127 (1991).

66. A. M. Spiegel, A. Shenker, and L. S. Weinstein, *Endocrinol. Rev.* **13**, 536–565 (1992).

67. E. J. Neer, *Protein Science*, **3**, 3–14 (1992).

68. J. H. Brown and D. Goldstein, *Mol. Pharmacol.* **30**, 566–570 (1986).

69. P. Samama and co-workers, *J. Biol. Chem.* **268**, 4625–4636 (1993).

70. Q. Ren and co-workers, *J. Biol. Chem.* **268**, 16,483–16,487 (1993).

71. M. S. Boguski and F. McCormick, *Nature*, **366**, 643–654 (1993).

72. S. E. Egan and R. A. Weinberg, *Nature*, **365**, 781–783 (1993).

73. U. Kikkawa, A. Kishimoto, and Y. Nishizuka, *Ann. Rev. Biochem.* **58**, 31–44 (1989).

74. B. Hille, *Ion Channels*, 2nd ed., Sinauer Association, Sunderland, Mass., 1992.

75. R. A. Janis, P. Silver, and D. J. Triggle, *Adv. Drug Res.* **16**, 309–591 (1987).

76. J.-L. Galzi and co-workers, *Ann. Rev. Pharmacol.* **31**, 37–72 (1991).

77. O. S. Anderson and R. E. Koeppe, *Physiol. Revs.* **72**, S89–S158 (1992).

78. D. J. Triggle and D. Rampe, *Drug Dev. Res.* **33**, 189–372 (1994).

79. A. Ullrich and J. Schlessinger, *Cell*, **61**, 203–212, 1990.

80. S. Raffioni, R. A. Bradshaw, and S. E. Busser, *Ann. Rev. Biochem.* **62**, 823–850 (1993).

81. J. Schlessinger and A. Ullrich, *Neuron*, **9**, 383–391 (1992).

82. N. D. Goldberg and M. K. Haddox, *Ann. Rev. Biochem.* **46**, 823–896 (1977).

83. D. L. Garbers, *Cell*, **71**, 1–4 (1992).

84. S. H. Snyder, *Science*, **257**, 494–496 (1992).

85. H. H. H. W. Schmidt, S. M. Lohmann, and U. Walter, *Biochem. Biophys. Acta*, **1178**, 153–175 (1993).

86. D. J. Triggle, *Trends Pharmacol. Sci.* **1**, 1–4 (1980).

87. G. B. Baker and A. J. Greenshaw, *Cell Mol. Neurobiol.* **9**, 1–44 (1989).

88. W. H. Frishman, *Am. J. Cardiol.* **59**, 26F–32F (1987).

89. O. Busto and co-workers, *New Eng. J. Med.* **315**, 854–859 (1986).

90. W. P. Hausdorff, M. G. Caron, and R. J. Lefkowitz, *FASEB J.* **4**, 2881 (1990).

91. R. L. Huganir and P. Greengard, *Neuron*, **5**, 555–567 (1990).

92. I. H. Pastan and M. C. Willingham, *Science*, **214**, 504–509 (1981).

93. J. L. Goldstein and co-workers, *Ann. Rev. Cell Biol.* **1**, 1–39 (1985).

94. I. Pastan and M. Willingham, *Endocytosis*, Plenum Press, New York, 1985.

95. E. Smythe and G. Warren, *Eur. J. Biochem.* **202**, 689–699 (1991).

96. R. J. Pollet and G. S. Levey, *Ann. Int. Med.* **92**, 663–680 (1980).

97. J. E. Fradkin and co-workers, *New Eng. J. Med.* **320**, 640–645 (1989).

98. A. M. Spiegel and co-workers, *New Eng. J. Med.* **312**, 26–33 (1985).

99. C. S. Scriver, ed., *The Metabolic Basis of Inherited Disease*, 6th ed., McGraw-Hill Book Co., Inc., New York, 1989.

100. D. Drachman, *New Eng. J. Med.* **330**, 1797–1810 (1994).

101. C. Peers and co-workers, *J. Physiol.* **421**, 293–308 (1990).

102. R. Llinas and co-workers, *Proc. Nat. Acad. Sci.* **90**, 11,743–11,747 (1993).

103. J. P. Kahn and co-workers, *Lancet*, 937–939 (Oct. 24, 1987).

104. R. J. Lefkowitz, M. G. Caron, and G. L. Stiles, *New Eng. J. Med.* **310**, 1570–1579 (1984).

105. H. H. Bode and J. D. Crawford, *New Eng. J. Med.* **280**, 750–754 (1969).

106. A. M. V. van der Ouweland and co-workers, *Nature Genetics*, **2**, 99–102 (1992).

107. Y. Pang and co-workers, *Nature Genetics*, **2**, 103–106 (1992).

108. P. Albright and co-workers, *Endocrinology*, **30**, 922–932 (1942).

109. M. A. Levine and co-workers, *Proc. Nat. Acad. Sci.* **85**, 617–621 (1988).

110. R. S. Weinstock and co-workers, *Nature*, **322**, 635–636 (1986).

111. A. G. Motulksy, *Science*, **231**, 126–129 (1986).

112. J. Davignon and M. Roy, *Am. J. Cardiol.* **72**, 6D–10D (1993).

113. M. J. Welsh and A. E. Smith, *Cell*, **73**, 1251–1254 (1993).

114. I. McIntosh and G. R. Cutting, *FASEB J.* **6**, 2775–2782 (1992).

115. S. Kingman, *J. NIH Res.* **7**, 34–35 (1995).

116. E. M. Sellers, in H. Kalant and W. H. E. Roshlau, *Principals of Medical Pharmacology*, B. C. Dekker, Inc., Toronto, Canada, 1985, pp. 109–118.

117. R. A. Lanius, B. A. Pasqualoto, and C. A. Shaw, *Trends Pharmacol. Sci.* **14**, 403–408 (1994).

118. W. Kalow and S. Spielberg, in Ref. 116, pp. 119–129.

119. J. Feely and T. Grimm, *Brit. J. Clin. Pharmacol.* **31**, 551–552 (1991).

120. H. H. Zhou and co-workers, *New Eng. J. Med.* **320**, 565–570 (1989).

121. B. J. Masterson, D. Reda, and W. G. Henderson, *New Eng. J. Med.* **328**, 914–921 (1993).

122. R. C. Parish and C. Spivey, *Brit. J. Clin. Pharmacol.* **31**, 197–199 (1991).

123. L. J. Notarianni, *Clin. Pharmacokin.* **18**, 20–36 (1990).

124. A. E. Reinberg, *Ann. Rev. Pharmacol. Toxicol.* **32**, 51–66 (1992).

General References

Refs. 1 and 2 are valuable discussions of drug action from the basic and clinical perspectives, respectively.

Ref. 4 presents a useful survey of the process of drug development from the regulatory perspective.

Ref. 6 is a classic exposition of the early development of drug–receptor interactions.

Ref. 12 is an excellent, although dated, source of information from the chemical perspective of drug–receptor interactions.

Ref. 44 by Kenakin is an excellent and detailed source of information for the quantitative analysis of drug–receptor interactions.

Ref. 74 is an outstanding volume dealing with ion channels and their behavior.

R. D. O'Brien, ed., *The Receptors*, Vol. 1, Plenum Press, New York, 1979.

David J. Triggle
State University of New York at Buffalo

PHARMACOKINETICS.

See PHARMACODYNAMICS.

PHASE EQUILIBRIUM.

See EXTRACTION; HIGH PRESSURE TECHNOLOGY; SEPARATIONS PROCESS SYNTHESIS.

PHENAZINE.

See ANTIOXIDANTS; ANTIOZONANTS; AZINE DYES; SULFUR DYES.

PHENAZINE ANTIBIOTICS.

See ANTIBIOTICS.

PHENAZONE.

See ANALGESICS, ANTIPYRETICS, AND ANTIINFLAMMATORY AGENTS; PYRAZOLES, PYRAZOLINES, AND PYRAZOLONES.

β-PHENETHYL ALCOHOL.

See BENZYL ALCOHOL AND β PHENYL ALCOHOLS; PERFUMES.

PHENETIDINES.

See ANALGESICS, ANTIPYRETICS, AND ANTIINFLAMMATORY AGENTS; XANTHENE DYES.

PHENOBARBITAL.

See HYPNOTICS, SEDATIVES, ANTICONVULSANTS, AND ANXIOLYTICS.

PHENOL

Phenol [108-95-2] is the common name of hydroxybenzene, C₆H₅OH, and belongs to the class of compounds, commonly referred to as phenols, containing one or more hydroxyl groups attached to an aromatic ring. Phenol has also been called carbolic acid, phenic acid, phenylic acid, phenyl hydroxide, or oxybenzene. The history of phenol goes back to 1834 when it was first isolated from coal tar and named carbolic acid. Until the advent of synthetic phenol production, just before World War I, coal tar remained the only source of phenol. The first synthetic phenol was produced by sulfonation of benzene and hydrolysis of the sulfonate. More than 99^o of phenol produced worldwide in the 1990s is from synthetic processes.

In 1993, worldwide phenol production was more than 5.2 million metric tons (1). The predominant uses of phenol are in phenolic resins (qv), bisphenol A, caprolactam (qv), aniline, and alkylphenols (qv).

Physical Properties

At room temperature phenol is a white, crystalline mass. Phenol gradually turns pink if it contains impurities or is exposed to heat or light. It has a distinctive sweet, tarry odor, and burning taste. Phenol has limited solubility in water between 0 and 65°C. Above 65.3°C phenol and water are miscible in all proportions. It is very soluble in alcohol, benzene, chloroform, ether, and partially disassociated organics in general. It is less soluble in paraffinic hydrocarbons. The important physical properties of phenol are listed in Table 1.

Table 1. Physical Properties of Phenol

Property	Value	Reference
molecular weight	94.11	2
boiling point at 101.3 kPa ^a	181.75	3

freezing point, °C	40.91	2
vapor pressure at 25°C, MPa ^a	46.84	4
flash point (closed cup), °C	79	2
density at 20°C (solid), g cm ³	1.0722	3
critical temperature, °C	421.1	3
critical pressure, MPa ^b	6.13	3
specific heat at 14–25°C, J gK ^c	2.35	5
heat of fusion, J g ^c	121.54	5
heat of vaporization, at bp, J g ^c	528.7	3
heat of combustion, kJ g ^c	32.468	3
viscosity, mPa·s(cP)		
at 50°C	3.49	3
70°C	2.03	3
90°C	1.26	3
specific heat		
at 4.0°C (solid)	1.24	6
227°C (solid)	1.41	6
70–74°C (liquid)	2.22	7

^a To convert kPa to mm Hg, multiply by 7.5.
^b To convert MPa to atm, divide by 0.1013.
^c To convert J to cal, divide by 4.184.

Chemical Properties

Phenol's chemical properties are characterized by the influences of the hydroxyl group and the aromatic ring upon each other. Although the structure of phenol is similar to cyclohexanol, phenol is a much stronger acid. Its pK_a in aqueous solution at 25°C is 9.89×10^{-10} (8). This characteristic allows aqueous hydroxides to convert phenol into their salts. The salts, especially those of sodium and potassium, are converted back into phenol by aqueous mineral acids or carboxylic acids.

Beside being acidic, a significant industrial chemical property of phenol is the extremely high reactivity of its ring toward electrophilic substitution. If steric conditions permit, the substitution leads first to the formation of the 2- or 4-monoderivative, then to the 2,4- or 2,6-diderivative, and finally to the 2,4,6-triderivative. The halogenation of phenol produces mono-, di-, and trihalophenols.

The most important commercial chemical reactions of phenol are condensation reactions. The condensation reaction between phenol and formaldehyde yields phenolic resins whereas the condensation of phenol and acetone yields bisphenol A (2,2-bis-(4-hydroxyphenol)propane). Phenolic resins and bisphenol A [80-05-7] account for more than two-thirds of U.S. phenol consumption (1).

Manufacture

The cumene oxidation route is the leading commercial process of synthetic phenol production, accounting for more than 95% of phenol produced in the world. The remainder of synthetic phenol is produced by the toluene oxidation route via benzoic acid. Other processes including benzene via cyclohexane, benzene sulfonation, benzene chlorination, and benzene oxychlorination have also been used in the manufacture of phenol. A list of U.S. phenol production plants and their estimated capacities in 1994 are shown in Table 2, and worldwide plants and capacities are shown in Table 3.

Table 2. Estimated U.S. Phenol Capacity,^a 1994

Company	Location	Capacity, 10 ³ t yr
Allied Chemical	Frankford, Pa.	367
Aristech Chemical	Haverhill, Ohio	296
BTL	Blue Island, Ill.	79
Dow Chemical	Oyster Creek, Tex.	250
Georgia Gulf	Plaquemine, La.	200
	Pasadena, Tex.	73
Mt. Vernon Phenol	Mt. Vernon, Ind.	290
Shell Chemical	Pasadena, Tex.	270
Texaco Chemical	El Dorado, Kan.	45
Total		1870

^a Cumene process; in addition, Kalama Co. (Kalama, Wash.) has a capacity of 32,000 t yr by the toluene process.

Table 3. Estimated World Synthetic Phenol Capacity,^a 1994

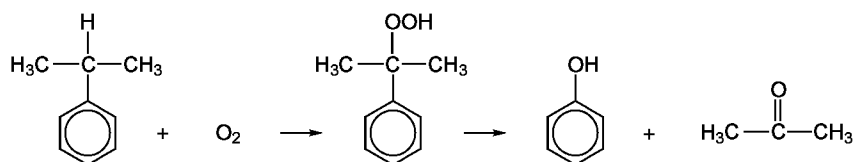
Region	Company	Location	Capacity, 10 ³ t yr
Mexico and South America	Rhodia	Sao Paulo, Brazil	115
	Fenoquimia	Veracruz, Mexico	40
total			155
Europe	Phenolchemie	Gladbeck, Germany	500
	Phenolchemie	Antwerp, Belgium	200
	Borealis	Porvoo, Finland	100
	Rhône Poulenc	Roussillon, France	150
	Leuna Werke	Leuna, Germany	55
	Enichem	Mantova, Italy	285
	Enichem	Porto Torres, Italy	100

	Ertisa	Huelva, Spain	120
	Neftochim	Burgas, Bulgaria	30
	Poland (state)	Pollock, Poland	45
	Petrochim	Brazi, Romania	75
	OKI	Zagreb, Croatia	7
	Grozny	Grozny, Russia	45
	Novukubyshevsk	Novukubyshevsk, Russia	60
	Omsk	Omsk, Russia	60
	OCK	Kazan, Russia	70
	Slovnaft	Bratislava, Czech	35
total			1967
Japan			
	Chiba Phenol	Chiba	200
	Mitsubishi Petrochemical	Kashima	180
	Mitsui Petrochemical	Chiba	220
	Mitsui Toatsu	Osaka	200
total			800
Asia			
	Kumho Shell Chemical	Korea	75
	Taiwan Prosperity Chem	Taiwan	100
	Huntsman Chemical	Australia	20
	Herdillia Chemical	Bombay, India	21
	Hindustan Organic Chem	Kerala, India	40
total			256
China			
	Yanshan P.C.	Beijing	60
	Shanghai Gaoqiao	Shanghai	57
	Shanghai P.C.	Jinshan	9
total			126
Worldwide total ^b			5174

^a Cumene process; in addition, DSM (Rotterdam, the Netherlands) and Shin Kippon (Kitakyushu, Japan) have a capacity of 32,000 and 120,000 t yr⁻¹, respectively, by the toluene process.

^b Including the United States.

Cumene Process. There are several licensed processes to produce phenol which are based on cumene (qv) (1,8–11). All of these processes consist of two fundamental chemical reactions: cumene is oxidized with air to form cumene hydroperoxide, and cumene hydroperoxide is cleaved to yield phenol and acetone. In this process, approximately 0.46 kg of acetone and 0.75 kg of phenol are produced per kg of cumene feedstock.



A typical phenol plant based on the cumene hydroperoxide process can be divided into two principal areas. In the reaction area, cumene, formed by alkylation of benzene and propylene, is oxidized to form cumene hydroperoxide (CHP). The cumene hydroperoxide is concentrated and cleaved to produce phenol and acetone. By-products of the oxidation reaction are acetophenone and dimethyl benzyl alcohol (DMBA). DMBA is dehydrated in the cleavage reaction to produce alpha-methylstyrene (AMS).

The recovery area of the plant employs fractionation to recover and purify the phenol and acetone products. Also in this section the alpha-methylstyrene is recovered and may be hydrogenated back to cumene or recovered as AMS product. The hydrogenated AMS is recycled as feedstock to the reaction area. The overall yield for the cumene process is 96 mol %. Figure 1 is a simplified process diagram.

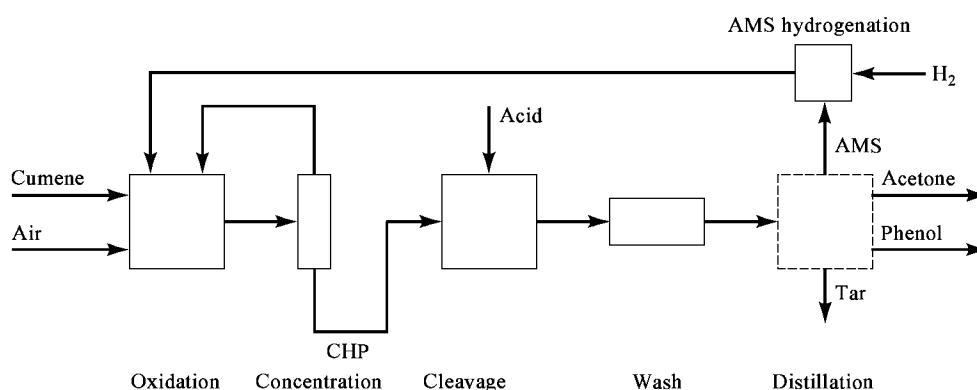


Fig. 1. Cumene process for phenol production.

Oxidation of cumene to cumene hydroperoxide is usually achieved in three to four oxidizers in series, where the fractional conversion is about the same for each reactor. Fresh cumene and recycled cumene are fed to the first reactor. Air is bubbled in at the bottom of the reactor and leaves at the top of each reactor. The oxidizers are operated at low to moderate pressure. Due to the exothermic nature of the oxidation reaction, heat is generated and must be removed by external cooling. A portion of cumene reacts to form dimethylbenzyl alcohol and acetophenone. Methanol is formed in the acetophenone reaction and is further oxidized to formaldehyde and formic acid. A small amount of water is also formed by the various reactions. The selectivity of the oxidation reaction is a function of oxidation conditions: temperature, conversion level, residence time, and oxygen partial pressure. Typical commercial yield of cumene hydroperoxide is about 95 mol % in the oxidizers. The reaction effluent is stripped off unreacted cumene which is then recycled as feedstock. Spent air from the oxidizers is treated to recover 99.99% of the cumene and other volatile organic compounds.

The concentrated cumene hydroperoxide solution from the cumene stripping section is fed to the cleavage reaction. The cleavage reaction is carried

out in the presence of an acid catalyst such as sulfuric acid. The cleavage reactor conditions are adjusted to maintain an optimum temperature to maximize phenol yield. Typical commercial yield of phenol from CHP in cleavage is greater than 98 mol %. The cleavage effluent contains the acid used as a catalyst for the cleavage reaction as well as formic and acetic acids which are by-products of the cleavage reactions. These must be neutralized and extracted to avoid corrosion problems downstream.

The neutralized cleavage product, consisting of acetone, phenol, water, hydrocarbons, and trace organic impurities, is separated in a series of distillation columns. Also in this section alpha-methylstyrene is either recovered as a product or hydrogenated to cumene.

Due to environmental considerations, many phenol plants are equipped with a special water treatment facility where acetone and phenol are recovered from the wastewater stream. Also, recovered heavy residue is considered a K-022 waste material by the U.S. EPA and must be properly disposed of by incineration or other means (12).

Safety is a critical aspect in the design of phenol plants. Oxidation of cumene to CHP occurs at conditions close to the flammable limits. Furthermore, the CHP is a potentially unstable material which can violently decompose under certain conditions. Thus, phenol plants must be carefully designed and provided with well-designed control and safety systems.

Toluene–Benzoic Acid Process. The toluene–benzoic acid process was first introduced by Dow-Canada, Ltd. in 1961 (13). It accounts for 4% of the total synthetic phenol capacity in the world.

The three chemical reactions in the toluene–benzoic acid process are oxidation of toluene to form benzoic acid, oxidation of benzoic acid to form phenyl benzoate, and hydrolysis of phenyl benzoate to form phenol. A typical process consists of two continuous steps (13,14). In the first step, the oxidation of toluene to benzoic acid is achieved with air and cobalt salt catalyst at a temperature between 121 and 177°C. The reactor is operated at 206 kPa gauge (2.1 kg cm² gauge) and the catalyst concentration is between 0.1 and 0.3%. The reactor effluent is distilled and the purified benzoic acid is collected. The overall yield of this process is believed to be about 68 mol % of toluene.

The second processing step, in which benzoic acid is oxidized and hydrolyzed to phenol, is carried out in two reactors in series. In the first reactor, the benzoic acid is oxidized to phenyl benzoate in the presence of air and a catalyst mixture of copper and magnesium salts. The reactor is operated at 234°C and 147 kPa gauge (1.5 kg cm² gauge). The phenyl benzoate is then hydrolyzed with steam in the second reactor to yield phenol and carbon dioxide. This occurs at 200°C and atmospheric pressure. The overall yield of phenol from benzoic acid is around 88 mol %. Figure 2 shows a simplified diagram for the toluene–benzoic acid process.

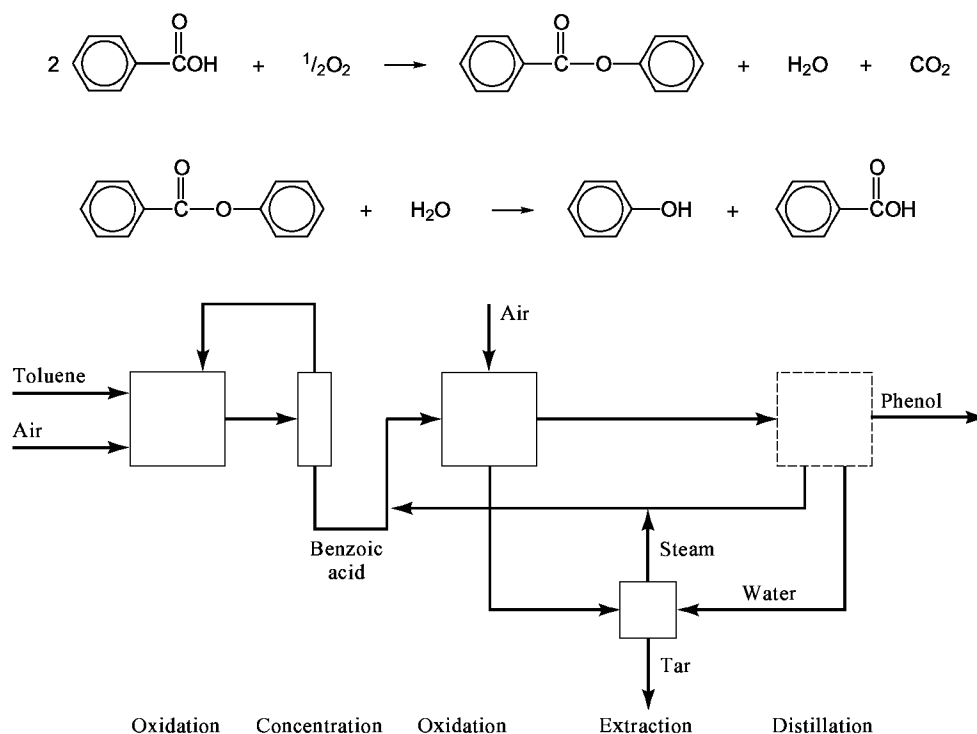


Fig. 2. Toluene–benzoic acid process for phenol production.

Other Processes.

Phenol Via Cyclohexene. In 1989 Mitsui Petrochemicals developed a process in which phenol was produced from cyclohexene. In this process, benzene is partially hydrogenated to cyclohexene in the presence of water and a ruthenium-containing catalyst. The cyclohexene then reacts with water to form cyclohexanol or oxygen to form cyclohexanone. The cyclohexanol or cyclohexanone is then dehydrogenated to phenol. No phenol plants have been built employing this process.

Benzenesulfonation. In the benzenesulfonation process, benzene reacts with concentrated sulfuric acid to form benzenesulfonic acid at about 150°C. The benzenesulfonic acid is neutralized with sodium sulfate to produce sodium benzenesulfonate, which is then fused with caustic soda to yield sodium phenate. The sodium phenate is acidified with sulfur dioxide and a small amount of sulfuric acid to release the phenol from the sodium salt. The phenol yield by this process can be as high as 88 mol % to that of the theoretical value based on benzene. Plants employing this technology have been shut down for environmental and economic reasons.

Benzenes Chlorination. In this process, benzene is chlorinated at 38–60°C in the presence of ferric chloride catalyst. The chlorobenzene is hydrolyzed with caustic soda at 400°C and 2.56 kPa (260 atm) to form sodium phenate. The impure sodium phenate reacts with hydrochloric acid to release the phenol from the sodium salt. The yield of phenol is about 82 mol % to that of the theoretical value based on benzene. Plants employing this technology have been shut down for environmental and economic reasons.

Benzenes Oxychlorination. In the benzenes oxychlorination process, also known as the Raschig Hooker process, benzene is oxychlorinated with hydrogen chloride, air, and with the presence of iron and copper chloride catalyst to form chlorobenzene. The reaction occurs at 200–260°C and atmospheric pressure. The chlorobenzene is hydrolyzed at 480°C in the presence of a suitable catalyst to produce phenol and chloride. The yield of phenol is ~90 mol% of theoretical. These plants have been shut down for environmental and economic reasons.

Economic Aspects

Table 4 shows the worldwide and U.S. production figures and prices for phenol since the mid-1980s. Because the cumene process accounts for more than 95% of the world's phenol supply, the economics of phenol production are closely tied to this production method. In the cumene process 615 kg of acetone are coproduced with each ton of phenol produced. Thus, the economics of phenol production are influenced by acetone (qv).

Table 4. Phenol Production and Prices

Year	Phenol consumption, t yr		U.S. price, ¢ kg
	U.S.	Worldwide	
1984	1310		
1985	1290		
1986	1413	3728	60
1987	1520	4007	99
1988	1615	4289	96
1989	1619	4495	101
1990	1605	4289	77
1991	1587	4352	58
1992	1694	4476	66
1993	1610	4477	70
1994 ^a	1646	4691	74

^a Values are estimated.

Worldwide, approximately 85° of acetone is produced as a coproduct with phenol. The remaining 17° is produced by on-purpose acetone processes such as the hydration of propylene to 2-propanol and the dehydrogenation of 2-propanol to acetone. The cost of production of 2-propanol sets the floor price of acetone as long as the acetone demand exceeds the coproduct acetone supply. However, there is a disparity in the growth rates of phenol and acetone, with phenol demand projected at 3.0° yr and acetone demand at 2.0° yr. If this continues, the coproduct supply of acetone will exceed the total acetone demand and on-purpose production of acetone will be forced to shut down; the price of acetone is expected to fall below the floor price set by the on-purpose cost production. Projections indicate that such a situation might occur in the world market by 2010. To forestall such a situation, companies such as Mitsui Petrochemical and Shinnippon (Nippon Steel) have built plants without the coproduction of acetone.

The key elements of the cost of production of phenol are feedstock cost and capital cost. For phenol produced on the U.S. Gulf Coast in a 200,000 t yr phenol plant built in 1994, the cumene feedstock cost represents 70° of the net cost of production, after allowance for acetone coproduct value. Depreciation of equipment represents 14° of the net cost and utilities approximately 7.6°. The remaining 8.4° covers all other expenses, including plant labor, maintenance, insurance, administration, sales, etc.

Specifications and Standards

DOT's Hazardous Materials Regulations (4,15) classifies phenol as a Class B poison. Drums and packages are to be labeled "Poison" and must comply with the regulations as contained in paragraph 173.369. Bulk containers must be properly placarded and marked. Shipping papers must include the required information regarding classification.

The U.S. *Pharmacopeia* (USP) specification for phenol includes (1) purity is to be no less than 98 wt °, (2) clear solubility of 1 part of phenol in 15 parts of water, (3) a congealing temperature to be not lower than 39°C, and (4) a content of nonvolatiles of no more than 0.05 wt °. Commercially, phenol specifications far exceed the USP requirement. Typical commercial phenol specifications are listed in Table 5. Higher purity material is required for some applications.

Table 5. Typical Commercial Phenol Specifications

Property	Value
freezing points, °C	40.85
color, ASTM Pt–Co	5
phenol concentration, wt °, min	99.99
water concentration, wt °, max	0.01
impurities, ppm	
carbonyls	40
hydroxyacetone	10
α-methylstyrene plus cumene	5
acetophenone	3
2-methylbenzofuran	2
Total (dry)	100

Analytical and Test Methods

Phenol quality tests and analyses can be divided into two categories: wet lab and liquid and gas chromatography. In the wet lab, phenol is tested for pH, solidification point, solubility in water, bromine index, color, and distillation ranges. Phenol concentration, impurities, and CHP contents are analyzed using highly automated liquid and gas chromatography.

In the wet lab, the minimum solidification points of freezing point is determined according to ASTM D1493-67. The color is determined according to ASTM D1686-61. The concentration of water in phenol is determined according to ASTM D1631.

Storage. Phenol is shipped in drums, tank trucks, and tank cars. It is loaded and shipped at elevated temperatures as a bulk liquid. In storage, phenol may acquire a yellow, pink, or brown discoloration which makes it unusable for some purposes. The discoloration is promoted by the action of water, light, air, and catalysts, eg, traces of iron or copper. When stored as a solid in the original drum or in nickel, glass-lined, or tanks lined with baked phenolic resin, phenol remains colorless for a number of weeks.

Storage tanks should be equipped with heating coils that pass upward through the entire vessel. Spillage and accidents have been caused by attempts to melt solid phenol using coils in the side of a tank which provide no escape for the expansion that occurs upon melting. Storage tanks may be constructed by either welding or riveting. Both horizontal and vertical tanks are suitable for phenol storage. Underground storage tanks are not recommended because leaks in the tanks are difficult to locate. Diking, draining, and tank support should be designed to conform with local regulations. Unloading and transfer should be operated so that there are no emissions into the atmosphere. Workmen should not be permitted to enter an empty tank which has been filled with phenol until it has been thoroughly cleaned.

Health and Safety Factors

Phenol fumes are irritating to the eyes, nose, and skin. According to the National Institute for Occupational Safety and Health (NIOSH), exposure to phenol should be controlled so that no employees are exposed to phenol concentrations >20 mg m⁻³, which is a time-weighted average concentration for up to a 10-h work day, 40-h work week. Phenol is very toxic to fish and has a nearly unique property of tainting the taste of fish if present in marine

environments at 0.1–1.0 ppm (7). Phenol presents no unusual fire hazard when handled at ambient temperatures, but burns if ignited. The lower flammability limit for the vapor is 1.5° in air. Phenol produces flammable, toxic vapors at elevated temperatures and has a flash point of 85°C (open cup).

Phenol is a general protoplasmic poison that is corrosive to any living tissue it contacts. It is a local anaesthetic, so that upon initial contact to the skin no pain is felt. By the time pain is felt, serious burns and absorption through skin may occur. Skin absorption occurs readily with a rapid onset of symptoms or death. Contact with eyes may cause severe damage and blindness. Upon contact with phenol, skin should be washed with water, then washed with poly(ethylene glycol) of molecular weight 300 (Macrogol 300) for at least 30 minutes. Eyes should be washed with flowing water for 10 minutes. Treatment for inhalation begins with removal of the exposed person to fresh air and provision of breathing support if needed.

Personnel who handle phenol should wear protective clothing, safety goggles, and rubber gloves, depending on the working conditions and amount of phenol handled.

Uses

The largest outlet for phenol worldwide is phenolic resins (qv). However, the growth rate of bisphenol A is higher than that of the other significant derivatives and is projected to become the principal use of phenol in the future (see EPOXY RESINS; POLYCARBONATES). Table 6 shows the portion of world phenol demand by use and the anticipated growth rate of the uses.

Table 6. World Demand for Phenol by Use, 1993

Phenol	Share, °	Growth rate, °
phenolic resins	35	2.2
bisphenol A	30	4.8
caprolactam	15	2.6
alkylphenols	7	2.6
aniline	5	4.2
others	8	1.5
Total ^a	100	

^a The average growth rate = 3.0.

Phenolic resins are produced by the condensation of phenol or a substituted phenol, such as cresol, with formaldehyde. These low cost resins have been produced commercially for more than 100 years and in the 1990s are produced by more than 40 companies in the United States. They are employed as adhesives in the plywood industry and in numerous under-the-hood applications in the automotive industry. Because of the cyclic nature of the automotive and home building industry, the consumption of phenol for the production of phenolic resins is subject to cyclic swings greater than that of the economy as a whole.

Bisphenol A (BPA), the fastest growing user of phenol, is produced by the condensation reaction of two moles of phenol and one mole of acetone. BPA, in turn, has two significant applications, which consume more than 80° of the production: polycarbonates, ie, engineering thermoplastics used for compact disks, ophthalmic lenses, automotive applications, and numerous other applications requiring the outstanding properties of polycarbonate; and epoxy resins, ie, thermosetting plastics employed in automotive coatings, electronic coatings, and other thermosetting applications. In markets such as Japan, BPA is already the leading use of phenol, whereas in other industrialized markets BPA is rapidly overtaking phenolic resins.

Some other phenol derivatives are somewhat local in application. For example, aniline is produced from phenol at only two plants, one in Japan and one in the United States. Likewise, phenol is used in the production of nylon, via caprolactam (qv) or adipic acid (qv) by only one United States producer and one European producer. These markets, like the phenolic resin and polycarbonate markets, are quite cyclical. Thus, the entire phenol market tends to be cyclical and closely tied to the housing and automotive markets.

BIBLIOGRAPHY

"Phenol" in *ECT* 1st ed., Vol. 10, pp. 279–297, by W. H. Prahl, Durez Plastics Inc.; in *ECT* 2nd ed., Vol. 15, pp. 147–159, by S. J. Lederman and N. Poffenberger; in *ECT* 3rd ed., Vol. 17, pp. 373–384, by C. Thurman, The Dow Chemical Co.

1. S. E. Howland and coworkers, "Phenol—A World Outlook," presented at the *1994 DeWitt Petrochemical Review*, Houston, Tex., 1994.
2. *The Merck Index*, 11th ed., Merck and Co., N.J., 1989.
3. *CRC Handbook of Chemistry and Physics*, 57th ed., CRC Press, Boca Raton, Fla., 1976–1977.
4. *Material Safety Data Sheets Collection*. Sheet No. 355, Revision C, Genium Publishing Corp., Schenectady, N.Y., Nov. 1990.
5. *Perry's Chemical Engineers' Handbook*, 6th ed., McGraw-Hill Book Co., Inc., New York, 1984.
6. R. J. L. Andan and co-workers, *Trans. Faraday Soc.* **59**, 830 (1963).
7. K. Verschueren, *Handbook of Environmental Data of Organic Compounds*, Van Nostrand Reinhold Co., New York, 1977, p. 525.
8. T. W. Graham-Solomons, *Organic Chemistry*, 5th ed., John Wiley & Sons, Inc., New York, 1992, p. 942.
9. *Hydrocarbon Proc.*, 155 (Nov. 1985).
10. *Hydrocarbon Proc.*, 168 (Mar. 1991).
11. P. R. Pujado, J. R. Salazar, and C. V. Berger, *Hydrocarbon Proc.*, 91–96 (Mar. 1976).
12. *Code of Federal Regulations*, Title 261.32, Part 40, U.S. Printing Office, Washington, D.C.
13. *Hydrocarbon Proc.* **44**(11) (Nov. 1965).
14. A. P. Gelbein and A. S. Nislick, *Hydrocarbon Proc.*, 125–128 (Nov. 1978).
15. *Code of Federal Regulations*, Title 49, U.S. Printing Office, Washington, D.C., pp. 100–199.

Jim Wallace
M. W. Kellogg Company

PHENOLIC FIBERS.

See PHENOLIC RESINS.

PHENOLIC RESINS

Phenolic resins are a large family of polymers and oligomers, composed of a wide variety of structures based on the reaction products of phenols with formaldehyde. Phenolic resins are employed in a wide range of applications, from commodity construction materials to high technology applications in electronics and aerospace. Generally, but not exclusively, thermosetting in nature, phenolic resins provide numerous challenges in the areas of synthesis, characterization, production, product development, and quality control.

As a family of resins originally developed in the early twentieth century, the nature and potential of phenolic resins have been explored thoroughly to produce an extensive body of technical literature (1–8). A symposium sponsored by the American Chemical Society commemorated 75 years of phenolic resin chemistry in 1983 (9), and in 1987 the Phenolic Molding Division of the Society of the Plastics Industry (SPI) sponsored a conference on phenolics in the twenty-first century (1).

Phenolic resins are prepared by the reaction of phenol or substituted phenol with an aldehyde, especially formaldehyde, in the presence of an acidic or basic catalyst. Their thermosetting character and the exotherm associated with the reaction presented technical barriers to commercialization. In 1900, the first U.S. patent was granted for a phenolic resin, using the resin in cast form as a substitute for hard rubber (10).

Work on the first commercially viable product was initiated by Baekeland in 1905. Using phenol and formaldehyde as starting materials, he established not only the differences between acid- and alkali-catalyzed products, but also the importance of excess phenol or formaldehyde made in producing intermediates. However, producing the resin was only part of the challenge. Baekeland also developed the technology to convert the reactive resins, which had a severe tendency to foam and cure to a brittle product, into useful molded articles by adding wood or mineral fibers and molding under heat and pressure. The final molded parts were tough, temperature resistant, and had a low void content (11). The first commercial phenolic resin plant was Bakelite GmbH, started in Germany in 1910; in the same year, the General Bakelite Co. was founded in the United States.

Early phenolic resins consisted of self-curing, resole-type products made with excess formaldehyde, and novolaks, which are thermoplastic in nature and require a hardener. The early products produced by General Bakelite were used in molded parts, insulating varnishes, laminated sheets, and industrial coatings. These areas still remain important applications, but have been joined by numerous others such as wood bonding, fiber bonding, and plywood adhesives. The number of producers in the 1990s is approximately 20 in the United States and over 60 worldwide.

Monomers

Phenol. This is the monomer or raw material used in the largest quantity to make phenolic resins (Table 1). As a solid having a low melting point, phenol, C₆H₅OH, is usually stored, handled in liquid form at 50–60°C, and stored under nitrogen blanket to prevent the formation of pink quinones. Iron contamination results in a black color.

Table 1. Properties of Phenol

Property	Value
mol wt	94.1
mp, °C	40.9
bp, °C	181.8
flash point, °C	79.0
autoignition temperature, °C	605.0
explosive limits, vol %	2–10
vapor pressure at 20°C, Pa ^a	20

^aTo convert Pa to mm Hg, multiply by 7.5 × 10^{–3}.

The most widely used process for the production of phenol is the cumene process developed and licensed in the United States by AlliedSignal (formerly Allied Chemical Corp.). Benzene is alkylated with propylene to produce cumene (isopropylbenzene), which is oxidized by air over a catalyst to produce cumene hydroperoxide (CHP). With acid catalysis, CHP undergoes controlled decomposition to produce phenol and acetone; α-methylstyrene and acetophenone are the by-products (12) (see CUMENE; PHENOL). Other commercial processes for making phenol include the Raschig process, using chlorobenzene as the starting material, and the toluene process, via a benzoic acid intermediate. In the United States, ~35–40% of the phenol produced is used for phenolic resins.

Substituted Phenols. Phenol itself is used in the largest volume, but substituted phenols are used for specialty resins (Table 2). Substituted phenols are typically alkylated phenols made from phenol and a corresponding α-olefin with acid catalysts (13). Acidic catalysis is frequently in the form of an ion-exchange resin (IER) and the reaction proceeds preferentially in the para position. For example, in the production of *t*-butylphenol using isobutylene, the product is >95% para-substituted. The incorporation of alkyl phenols into the resin reduces reactivity, hardness, cross-link density, and color formation, but increases solubility in nonpolar solvents, flexibility, and compatibility with natural oils.

Table 2. Substituted Phenols Used for Phenolic Resins

Substituted phenol	Resin application
cresol (<i>o</i> -, <i>m</i> -, <i>p</i> -)	coatings, epoxy hardeners
<i>p</i> - <i>t</i> -butylphenol	coatings, adhesives
<i>p</i> -octylphenol	carbonless paper, coatings
<i>p</i> -nonylphenol	carbonless paper, coatings
<i>p</i> -phenylphenol	carbonless paper

bisphenol A	low color molding compounds, coatings
resorcinol	adhesives
cashew nutshell liquid	friction particles

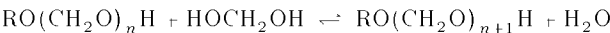
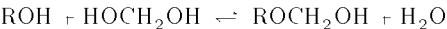
Formaldehyde. In one form or another, formaldehyde is used almost exclusively in the production of phenolic resins, regardless of the type of phenol (Table 3). It is frequently produced near the site of the resin plant by either of two common processes using methanol (qv) as the raw material. In the silver catalyst process, the reaction takes place at 600–650°C and produces water and hydrogen as by-products. The more common metal oxide process operates at 300–400°C. The gaseous formaldehyde is absorbed in water, the final product is a formalin solution containing 36–50% formaldehyde. Of the various chemical forms of formaldehyde, the aqueous form is preferred for making phenolic resins, even though at least half of this form is water. The water serves to moderate the reaction and is readily removed in processing equipment (14) (see FORMALDEHYDE).

Table 3. Forms of Formaldehyde

Type	Chemical formula	Resin preparation	
		Advantages	Disadvantages
gaseous	CH ₂ O		unstable
formalin			
36% _o	HO(CH ₂ O) _n H ^a	easy handling, moderate reactivity, stable at RT	high water content
50% _o	HO(CH ₂ O) _n H ^b	increased capacity	elevated temperature storage, formic acid formation
paraformaldehyde	HO(CH ₂ O) _n H ^c	increased capacity, water-free	dangerously high reactivity, solids handling
trioxane	(CH ₂ O) ₃	water-free	catalyst requirements, high cost
hexamethylene-tetramine	(CH ₂) ₆ N ₄	autocatalytic	amine incorporation

^a n ≈ 2.
^b n ≈ 3.
^c n = 20–100.

Aqueous Formaldehyde. Water solutions of formaldehyde consist mainly of telomers of methylene glycol having <100 ppm of the formaldehyde as CH₂O (5). Alcohols form hemiformals with aqueous formaldehyde according to the following, where n = 1, 2, 3, etc.



However, a second mole of alcohol or hemiformal cannot be added at the ordinary pH of such solutions. The equilibrium constant for hemiformal formation depends on the nature of the R group of the alcohol. Using nmr spectroscopy, a group of alcohols including phenol has been examined in solution with formaldehyde (15,16). The spectra indicated the degree of hemiformal formation in the order of <methanol> benzyl alcohol >phenol. Hemiformal formation provides the mechanism of stabilization; methanol is much more effective than phenol in this regard.

The large value for the hemiformal formation constant of methanol and its low molecular weight explains the high efficiency of methanol in stabilizing formalin solutions. Phenol, on the other hand, is inefficient, and phenol hemiformals are only formed by careful removal of water (17).

Other Aldehydes. The higher aldehydes react with phenol in much the same manner as formaldehyde, although at much lower rates. Examples include acetaldehyde, CH₃CHO; paraldehyde, (CH₃CHO)₃; glyoxal, OCH–CHO; and furfural. The reaction is usually kept on the acid side to minimize aldol formation. Furfural resins, however, are prepared with alkaline catalysts because furfural self-condenses under acid conditions to form a gel.

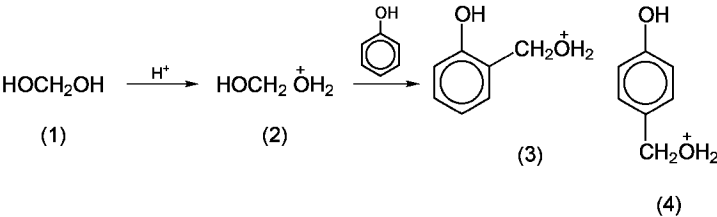
Hexamethylenetetramine. Hexa, a complex molecule with an adamantane-type structure, is prepared from formaldehyde and ammonia, and can be considered a latent source of formaldehyde. When used either as a catalyst or a curative, hexa contributes formaldehyde-residue-type units as well as benzylamines. Hexa [100-97-0] is an infusible powder that decomposes and sublims above 275°C. It is highly soluble in water, up to ca 45 wt % with a small negative temperature solubility coefficient. The aqueous solutions are mildly alkaline at pH 8–8.5 and reasonably stable to reverse hydrolysis.

Other Reactants. Other reactants are used in smaller amounts to provide phenolic resins that have specific properties, especially coatings applications. Aniline had been incorporated into both resoles and novolaks but this practice has been generally discontinued because of the toxicity of aromatic amines. Other materials include rosin (abietic acid), dicyclopentadiene, unsaturated oils such as tung oil and linseed oil, and polyvalent cations for cross-linking.

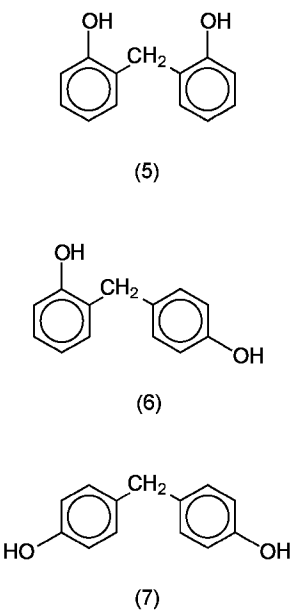
Polymerization

Phenolic resins are prepared with strong acid or alkaline catalysts. Occasionally, weak or Lewis acids, such as zinc acetate, are used for specialty resins.

Strong-Acid Catalysts, Novolak Resins. Phenolic novolaks are thermoplastic resins having a molecular weight of 500–5000 and a glass-transition temperature, T_g of 45–70°C. The phenol–formaldehyde reactions are carried to their energetic completion, allowing isolation of the resin; formaldehyde–phenol molar ratios are between 0.5:1 and 0.8:1. Methylene glycol [463-57-0] (1) is converted to the corresponding hydrated carbonium ion (2), which adds to the ortho and para positions of phenol with the elimination of water to form the corresponding ortho (3) and para (4) benzylic ions. The benzylic carbonium ions are in equilibrium with the corresponding benzylic alcohols, observed by nmr as transient species in the formation of novolak resins (15).



In the next step the hydrated benzylic carbonium ions (3 and 4) react with free ortho and para positions on phenols to form methylene-linked bisphenols, 2,2' (5), 2,4' (6), and 4,4' (7).



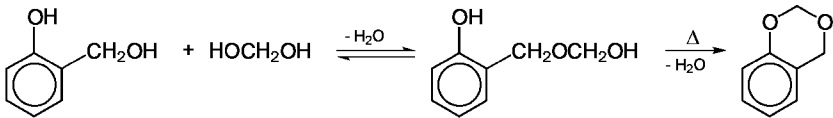
Continued reaction leads to the formation of novolak polymers having a molecular weight of up to 5000. Acid-catalyzed resins contain 50–75% 2,4' linkages (6). The reaction rate is proportional to catalyst, formaldehyde, and phenol concentrations, and inversely proportional to the concentration of water. The rate of formation of the benzyl alcohol intermediate is 5–10 times lower than the rate to form the methylene-linked bisphenol (3). At typical molecular weights of 500–1000, novolak molecules are essentially linear because of the much lower reactivity of doubly reacted phenolic units. In higher molecular weight polymers, the low concentration of end groups and unreacted phenol causes branching. Above 1000 in molecular weight, branching has been observed by ¹³C-nmr; about 20% branching has been predicted in computer simulations (13,18,19). In the curing process, end groups are more reactive than the backbone groups. Thus a branched resin having a higher content of end groups than a corresponding linear equivalent may gel sooner and cure faster because of the higher resin functionality. The properties of an acid-catalyzed phenolic resin are shown in Table 4.

Table 4. Novolak Resin Properties

Property	Catalyst	
	Acid	Zn acetate ^a
formaldehyde–phenol molar ratio	0.75	0.60
nmr analysis, %		
2,2'	6	45
2,4'	73	45
4,4'	21	10
gpc analysis		
phenol, %	4	7
<i>M_n</i>	900	550
<i>M_w</i>	7300	1800
water, %	1.1	1.9
<i>T_g</i> , °C	65	48
gel time, s	75	25

^a High ortho.

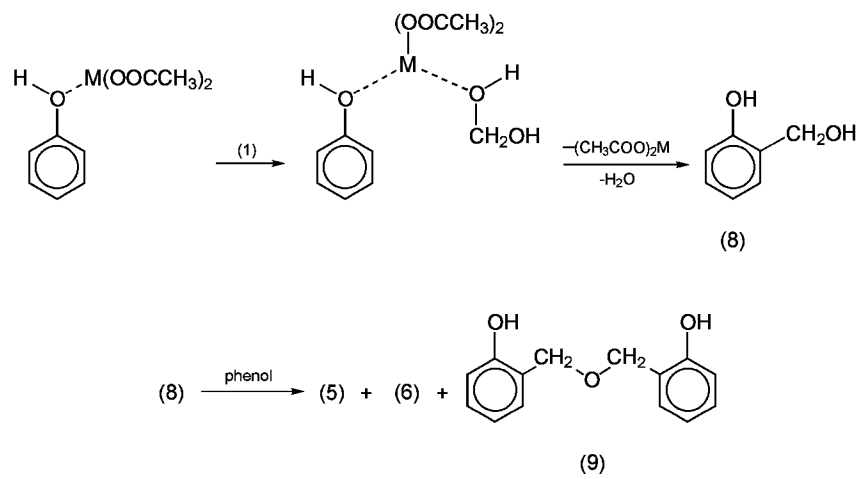
The typical acid catalysts used for novolak resins are sulfuric acid, sulfonic acid, oxalic acid, or occasionally phosphoric acid. Hydrochloric acid, although once widely used, has been abandoned because of the possible formation of toxic chloromethyl ether by-products. The type of acid catalyst used and reaction conditions affect resin structure and properties. For example, oxalic acid, used for resins chosen for electrical applications, decomposes into volatile by-products at elevated processing temperatures. Oxalic acid-catalyzed novolaks contain small amounts (1–2% of the original formaldehyde) of benzodioxanes formed by the cyclization and dehydration of the benzyl alcohol hemiformal intermediates.



Benzodioxane is reasonably stable at neutral pH, but may decompose when the resin is cured, serving as a source of labile formaldehyde. Benzodioxanes are not found in sulfuric or sulfonic acid-catalyzed resins, since the stronger acid readily catalyzes the second step in the reaction sequence.

Neutral Catalysts, High Ortho Novolaks. In the range of pH 4–7, formaldehyde substitution of the phenolic ring is possible, using divalent metal catalysts containing Zn, Mg, Mn, Cd, Co, Pb, Cu, and Ni; certain aluminum salts are also effective. Organic carboxylates are required as anions in order to obtain sufficient solubility of the catalyst in the reaction medium, as well as to provide a weak base. Acetates are most convenient and economical. Although lead acetate is highly effective because of its excellent solubility properties, it has been largely eliminated because of lead toxicity. Zinc and calcium salts are probably the most widely used catalysts (20).

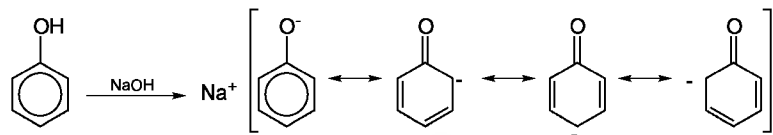
Novolaks produced from these catalysts exhibit a high content of 2,2'-methylene units. The mechanism proposed for the ortho-directing effect involves chelation of the phenolic unit with the metal ion.



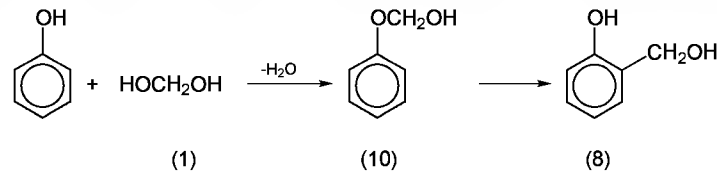
Zinc acetate catalyst produces essentially 100% *o*-methylol phenol (8) in the first step. The second step gives an approximately equal quantity of 2,2'-(5, 45%) and 2,4'-diphenylmethylene (6, 45%) bridges, indicating little chelate-directing influence. In addition, a small quantity (10%) of methylene ether units (9) (dibenzyl ether) is observed at moderate reaction temperature.

High ortho novolaks have faster cure rates with hexa. Typical properties of a zinc acetate-catalyzed high ortho novolak are also shown in Table 4. The gel time with hexa is one-third of that with a strong acid-catalyzed novolak.

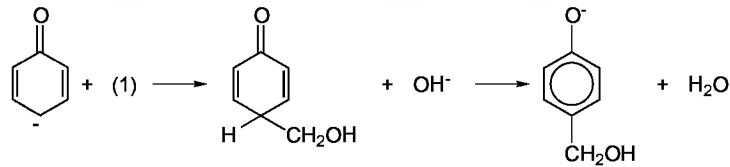
Alkaline Catalysts, Resoles. Resole-type phenolic resins are produced with a molar ratio of formaldehyde to phenol of 1.2:1 to 3.0:1. For substituted phenols, the ratio is usually 1.2:1 to 1.8:1. Common alkaline catalysts are NaOH, $\text{Ca}(\text{OH})_2$, and $\text{Ba}(\text{OH})_2$. Whereas novolak resins and strong acid catalysis result in a limited number of structures and properties, resoles cover a much wider spectrum. Resoles may be solids or liquids, water-soluble or -insoluble, alkaline or neutral, slowly curing or highly reactive. In the first step, the phenolate anion is formed by delocalization of the negative charge to the ortho and para positions.



Alkaline catalysts are also effective in the polymerization–depolymerization of methylene glycol. The mechanism of the formaldehyde addition to the phenolate is still not completely understood. The most likely mechanism involves the contribution of phenol hemiformals (10) (5).



Rate studies show that base-catalyzed reactions are second order and depend on the phenolate and methylene glycol concentrations. The most likely path involves a nucleophilic displacement by the phenoxide on the methylene glycol (1), with the hydroxyl as the leaving group. In alkaline media, the methylolated quinone intermediate is readily converted to the phenoxide by hydrogen-ion abstraction (21).



The ratio of ortho-to-para substitution depends on the nature of the cation and the pH. Para substitution is favored by K^+ and Na^+ ions and higher pH, whereas ortho substitution is favored at lower pH and by divalent cations, such as Ba^{2+} , Ca^{2+} , and Mg^{2+} (22).

Several extensive kinetic studies on the polymethylation of phenol have been reported (21,23,24). For the reaction scheme shown in Figure 1, seven different rate constants must be determined. Despite different solution concentration, temperatures, and methods of analysis, comparing reaction rates (25–27) from each study using an NaOH catalyst gave fairly close agreement that rate constants increase with methylol substitution. In fact, dimethylol-substituted phenols react with formaldehyde two to four times faster than phenol. As a result, unreacted phenol remains high in resole resins (5–15%) even though the formaldehyde:phenol ratio is as high as 3:1.

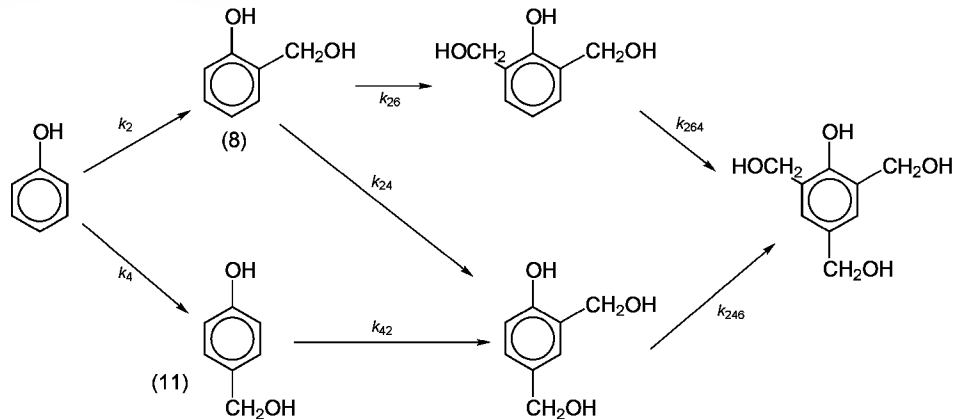


Fig. 1. Possible pathways and rate constants for the methylation of phenol.

4-CH ₂ OH	12	9
4-CH ₂ OCH ₂ OH	16	0
4-CH ₂ OR	2	4
2,2'-CH ₂	0	0
2,4'-CH ₂	7	12
4,4'-CH ₂	7	10
2-CH ₂ N	0	27
4-CH ₂ N	0	7
benzoxazine	0	2

^a 6 pph.

Manufacture

The final state of a phenolic resin varies dramatically from thermoplastic to thermoset and from solid to liquid, and includes solutions and dispersions. With a bulk process, resole resins, in neat or concentrated form, must be produced in small batches (ca 2–9.5 m³) in order to maintain control of the reaction and obtain a uniform product. On the other hand, if the product contains a large amount of water, such as liquid plywood adhesives, large reactors (19 m³) can be used. Melt-stable products such as novolaks can be prepared in large batches (19–38 m³) if the exotherms can be controlled.

Batch processes for most phenolic resins employ the equipment shown in Figure 2. Liquid reactants are metered into the stirred reaction vessel through weigh tanks, whereas solid reactants such as bisphenol A and Ba(OH)₂ present handling problems. Facilities are provided to carry out the reaction under a vacuum or an inert gas.

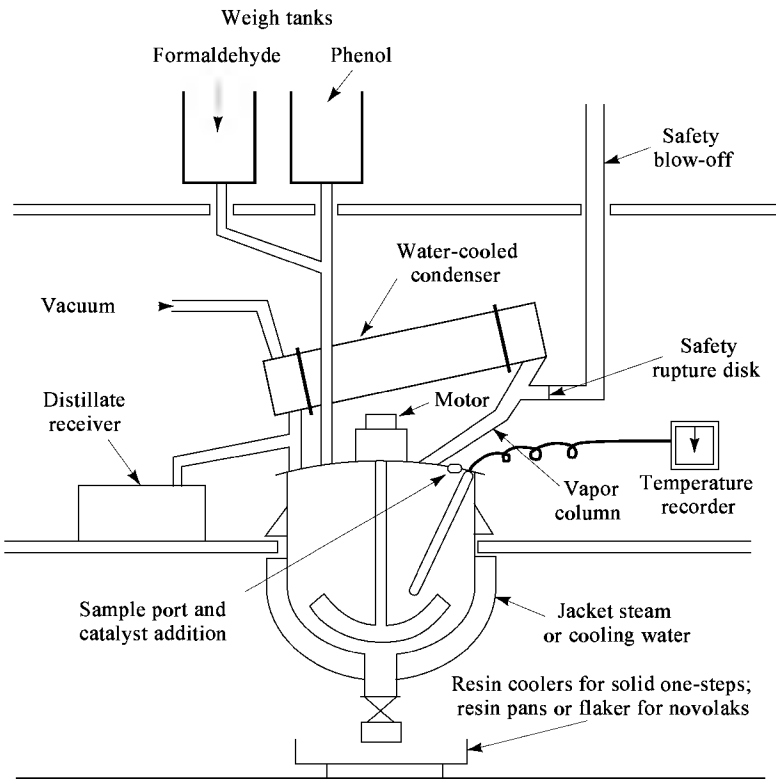


Fig. 2. Typical phenolic resin production unit.

Materials of Construction. Compatibility of the materials of construction and the process chemicals is extremely important. The reactors are usually made of stainless steel alloys. Copper is avoided because of the possible presence of amines. Glass-lined reactors are occasionally used for nonalkaline resins. Because the use of HCl has been largely discontinued, material requirements are less stringent. The reactor contains a bottom discharge, which for solid heat-reactive resins must be large. Solid resole resins are discharged for rapid cooling in order to quench the thermosetting reactions. Resin coolers are made up of vertical plates with internally circulating water. The product can also be discharged to a large cooled surface. Discharges to belt and drum flakers are highly automated, however they can only be used for less-reactive resins.

Novolak resins can be stored molten in heated holding tanks under nitrogen. Because novolaks are used mainly in pulverized form with hexa and additives, a process that includes belt flaking and feeding directly into the blending and pulverizing system is preferred. Liquid and solution resole resins are cooled in the reactor by using jacket cooling and vacuum refluxing. Discharged products are filtered and pumped to refrigerated intermediate holding areas or packaged for shipping. The stability of liquid resole products varies greatly from product to product and depends on the storage temperature. The viscosity of a liquid resole resin increases but the water miscibility decreases as time and temperature increase. Generally, resoles, both liquids and solids, must be refrigerated.

Novolak Resins. In a conventional novolak process, molten phenol is placed into the reactor, followed by a precise amount of acid catalyst. The formaldehyde solution is added at a temperature near 90°C and a formaldehyde-to-phenol molar ratio of 0.75:1 to 0.85:1. For safety reasons, slow continuous or stepwise addition of formaldehyde is preferred over adding the entire charge at once. Reaction enthalpy has been reported to be above 80 kJ/mol (19 kcal/mol) (29,30). The heat of reaction is removed by refluxing the water combined with the formaldehyde or by using a small amount of a volatile solvent such as toluene. Toluene and xylene are used for azeotropic distillation. Following decantation, the toluene or xylene is returned to the reactor.

The reaction is completed after 6–8 h at 95°C; volatiles, water, and some free phenol are removed by vacuum stripping up to 140–170°C. For resins requiring phenol in only trace amounts, such as epoxy hardeners, steam distillation or steam stripping may be used. Both water and free phenol affect the cure and final resin properties, which are monitored in routine quality control testing by gc. Oxalic acid (1–2 parts per 100 parts phenol) does not require neutralization because it decomposes to CO, CO₂, and water; furthermore, it produces milder reactions and low color. Sulfuric and sulfonic acids are strong catalysts and require neutralization with lime; 0.1 parts of sulfuric acid per 100 parts of phenol are used. A continuous process for novolak resin production has been described (31,32). An alternative process for making novolaks without acid catalysis has also been reported (33), which uses a

peroxidase enzyme to polymerize phenols in an aqueous solution. The enzyme can be derived from soybeans or horseradish.

High Ortho Novolaks. The process for high ortho novolaks is similar to the one used for those catalyzed by strong acid. Zinc acetate is used at concentrations higher than the acids, typically 2° or more. The formaldehyde:phenol ratio is similar (0.75–0.85) but yields are 5–10° lower than those produced with strong acids, and reaction times are longer. Problems with gel particles and bulk gelation occur more frequently because small amounts of reactive dibenzyl ether groups are present. Overall, the process is more expensive because of higher raw material costs, lower yields, and longer cycle times.

Another process employs a pH maintained at 4–7 and a catalyst that combines a divalent metal cation and an acid. Water is removed continuously by azeotropic distillation and xylene is recycled. The low water content increases the reaction rate. The dibenzyl ether groups are decomposed by the acid; the yield of 2,2'-methylene can be as high as 97° (34).

Resoles. Like the novolak processes, a typical resole process consists of reaction, dehydration, and finishing. Phenol and formaldehyde solution are added all at once to the reactor at a molar ratio of formaldehyde to phenol of 1.2–3.0:1. Catalyst is added and the pH is checked and adjusted if necessary. The catalyst concentration can range from 1–5° for NaOH, 3–6° for Ba(OH)₂, and 6–12° for hexa. A reaction temperature of 80–95°C is used with vacuum-reflux control. The high concentration of water and lower enthalpy compared to novolaks allows better exotherm control. In the reaction phase, the temperature is held at 80–90°C and vacuum-refluxing lasts from 1–3 h as determined in the development phase. Solid resins and certain liquid resins are dehydrated as quickly as possible to prevent overreacting or gelation. The end point is found by manual determination of a specific hot-plate gel time, which decreases as the polymerization advances. Automation includes on-line viscosity measurement, gc, and gpc.

Phenolic Dispersions. These systems are predominantly resin-in-water systems in which the resin exists as discrete particles. Particle size ranges from 0.1 to 2 µm for stable dispersions and up to 100 µm for dispersions requiring constant agitation. Some of the earliest nonaqueous dispersions were developed for coatings applications. These systems consist of an oil-modified phenolic resin complexed with a metal oxide and a weak solvent.

In the post-dispersion process, the solid phenolic resin is added to a mixture of water, cosolvent, and dispersant at high shear mixing, possibly with heating. The cosolvent, frequently an alcohol or glycol ether, and heat soften the resin and permit small particles to form. On cooling, the resin particles, stabilized by dispersant and perhaps thickener, harden and resist settling and agglomeration. Both resole and novolak resins have been made by this process (25).

The *in situ* process is simpler because it requires less material handling (35); however, this process has been used only for resole resins. When phenol is used, the reaction system is initially one-phase; alkylated phenols and bisphenol A present special problems. As the reaction with formaldehyde progresses at 80–100°C, the resin becomes water-insoluble and phase separation takes place. Catalysts such as hexa produce an early phase separation, whereas NaOH-based resins retain water solubility to a higher molecular weight. If the reaction medium contains a protective colloid at phase separation, a resin-in-water dispersion forms. Alternatively, the protective colloid can be added later in the reaction sequence, in which case the reaction mass may temporarily be a water-in-resin dispersion. The protective colloid serves to assist particle formation and stabilizes the final particles against coalescence. Some examples of protective colloids are poly(vinyl alcohol), gum arabic, and hydroxyethylcellulose.

For products intended to remain stable dispersions for an extended period, a particle size of 2 µm or less is desirable. A thickening agent is usually added after the reaction has been completed and the mixture is cooled in order to prevent settling and agglomeration. Examples of thickeners are guar gum, xanthan gum, and hydroxyethylcellulose. The final products are generally between 40 and 50° solids, with a viscosity of 1500–5000 mPas (cP).

Resole dispersions intended for isolation as discrete particles (26) can be used as flatting agents in coatings (27). Particles larger than 1000 µm are used in friction-element compositions. A-stage, thermosetting phenolic particles have been isolated from dispersion (26,36). With a hexa catalyst (6–12 parts) and a formaldehyde:phenol ratio of 1.5:1, the reaction is carried out at 50° solids for ca 90 min at 85°C. Poly(vinyl alcohol) and gum arabic are the preferred protective colloids. The particles (20–80 µm) are isolated from the mixture by filtration and, in the patent examples, by fluid-bed drying. These A-stage products (gel time at 150°C, 50–100 s) are suitable in applications where pulverized phenolic resins are being used, as well as in applications that take advantage of their spherical nature. One patent describes a sinter-resistant product for wood-bonding applications (37). In another patented process, both the production of particulate novolak resins and the aqueous dispersions of these resins are described (38).

Spray-Dried Resins. Spray drying produces resins in particulate form. Spray-drying a resole solution containing a blowing agent (39) produces phenolic microballoons. Spray drying also produces A-stage resins (40). The resins, prepared with a high NaOH content, are spray dried to give a final particle size of 40–60 µm. The particles are hygroscopic because of the high caustic content, but are sinter-resistant when kept dry. The principal application for this type of product is believed to be wood binding, especially for waferboard applications.

Cure

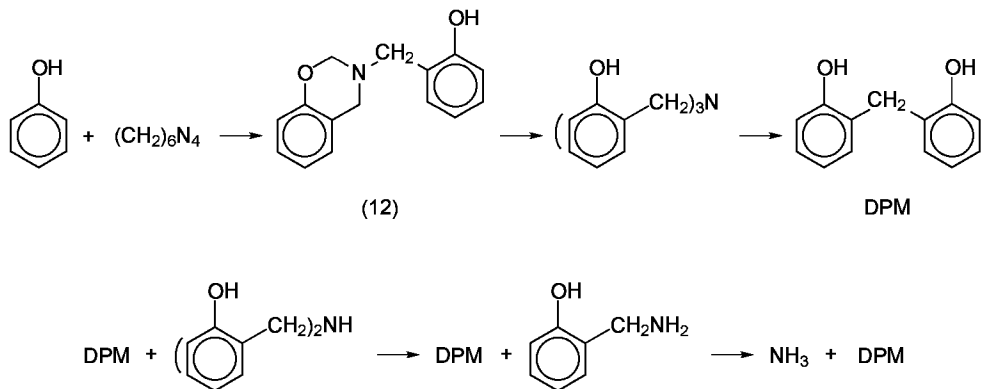
A typical resin has an initial molecular weight of 150 to perhaps 1500. For systems of unsubstituted phenols, the final cross-link density is 150–300 atomic mass units (amu) per cross-link. In other words, 25–75° of the ring-joining reactions occur during the cure phase.

Resoles. The advancement and cure of resole resins follow reaction steps similar to those used for resin preparation; the pH is 9 or higher and reaction temperature should not exceed 180°C. Methylol groups condense with other methylols to give dibenzyl ethers and react at the ortho and para positions on the phenol to give diphenylmethylenes. In addition, dibenzyl ethers eliminate formaldehyde to give diphenylmethanes.

In some resole applications, such as foam and foundry binders, a rapid cure of a liquid resin is obtained at RT with strong acid. The reactions proceed in the same manner as those of novolak resin formation. Methylol groups react at ortho and para phenolic hydrogen to give diphenylmethane units (41).

At pH 4–6, the cure is slower than it is at pH 8 and higher, and much slower than at pH 1–3. Reactions at pH 4–6 resemble those on the more alkaline side, but with a substantial increase in side-products. This is partly the result of the low rates of the main reactions and partly the result of stable intermediates at this pH range.

Novolaks. Novolak resins are typically cured with 5–15° hexa as the cross-linking agent. The reaction mechanism and reactive intermediates have been studied by classical chemical techniques (3,4) and the results showed that as much as 75° of nitrogen is chemically bound. More recent studies of resin cure (42–45) have made use of tga, dta, gc, ir, and nmr (15). They confirm that the cure begins with the formation of benzoxazine (12), progresses through a benzyl amine intermediate, and finally forms (hydroxy)diphenylmethanes (DPM).



In the reaction of phenol and bisphenol F with hexa, nmr spectra show the transient appearance of benzoxazine intermediates; after 2 h at 103°C, all the benzoxazine decomposed to the diphenylmethylene and benzylamine intermediates (15).

The cure of novolaks with hexa has been studied with differential scanning calorimetry (dsc) and torsional braid analysis (tba) (46); both a high ortho novolak and a conventional acid-catalyzed system were included. The dsc showed an exothermic peak indicating a novolak–hexa reaction ca 20°C higher than the gelation peak observed in tba. Activation energies were also calculated.

Property	Acid-catalyzed	High ortho
tba gel temperature, °C	130	113
dsc exotherms, °C	150	138

The resin rich in 2,2'-methylene exhibited the lowest activation energy, gel temperature, and dsc exotherm. The high concentration of the slightly acidic 2,2'-diphenylmethane end groups may account for the higher reactivity. These end groups should react with hexa to form benzoxazine intermediates first, which then decompose to react with vacant positions throughout the novolak molecule.

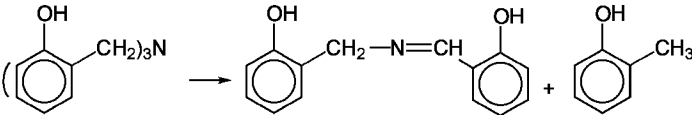
An isothermal method for studying the cure of phenolics employs dynamic mechanical analysis (dma) (Table 7). The problems associated with programmed heating rates are avoided and mathematical treatment of the results is simplified. Although a more complex treatment is possible, a simple first-order dependence of modulus with time and an Arrhenius-type temperature dependence are sufficient. The rate studies of Table 7 indicate that doubling the amount of hexa doubles the rate at which the modulus approaches its long-term value. The novolak–12° hexa cures substantially slower than the resole. In addition, they differ in temperature dependence of cure rates; the resole has an activation energy approximately four times greater than that of the novolak–12° hexa (47).

Table 7. Isothermal Dma Results

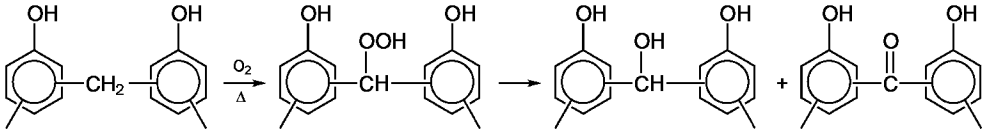
Sample	Rate, min ⁻¹		Activation energy, kJ mola
	at 150°C	at 185°C	
hexa-catalyzed resole	0.22	1.00	71.1
novolak			
6° hexa	0.07	0.09	8.8
12° hexa	0.12	0.19	18.8

^a To convert kJ to kcal, divide by 4.184.

Decomposition of Cured Resoles and Novolaks. Above 250°C, cured phenolic resins begin to decompose. For example, dibenzyl ethers such as (9) disproportionate to aldehydes (salicylaldehyde) and cresols (*o*-cresol). The aldehyde group is rapidly oxidized to the corresponding carboxylic acid. In an analogous reaction in hexa-cured novolaks, tribenzylamines decompose into cresols and azomethines, which cause yellowing.



Substantial decomposition of phenolic resins begins above 300°C. In the presence of oxygen, the methylene bridging group is converted to a hydroperoxide which in turn yields alcohols and ketones on decomposition.



The ketone is especially susceptible to random chain scission. Decomposition continues up to ca 600°C; the by-products are mostly water, CO, CO₂, and phenols. The first stage of decomposition produces a porous structure having minimal shrinkage. The second stage begins near 600°C and is accompanied by shrinkage and substantial evolution of CO₂, H₂O, methane, and aromatics. The resulting polyaromatic chars represent ca 60° of the original resin when the atmosphere is inert, but this may be substantially less in the presence of air. The char ignites in air above 900°C.

Analysis and Characterization

The principal techniques for determining the microstructure of phenolic resins include mass spectroscopy, proton, and ¹³C-nmr spectroscopy, as well as gc, lc, and gpc. The softening and curing processes of phenolic resins are effectively studied by using thermal and mechanical techniques, such as tga, dsc, and dynamic mechanical analysis (dma). Infrared (ir) and electron spectroscopy are also employed.

Spectroscopy. Infrared spectroscopy (48) permits structural definition, eg, it resolves the 2,2'- from the 2,4'-methylene units in novolak resins. However, the broad bands and severely overlapping peaks present problems. For uncured resins, nmr rather than ir spectroscopy has become the technique of choice for microstructural information. However, Fourier transform infrared (ftir) gives useful information on curing phenolics (49). Nevertheless, ir spectroscopy continues to be used as one of the detectors in the analysis of phenolics by gpc.

The first magnetic resonance technique that provided structural information on phenolic resins was proton magnetic resonance (pmr). Even given the low degree of chemical-shift sensitivity, substantial microstructural information has been obtained on both novolaks and resoles. Most of the information has arisen from resolving and assigning peaks in the CH₂ region of the spectrum, at chemical shifts ranging from 3.5–5.5 ppm in pyridine solvent (50–52). A variety of methylene units, including methylols and benzylamines, can be resolved in resoles by examining the spectrum before and after acetylation with acetic anhydride. Table 8 gives the methylene structures and chemical-shift assignments for phenolic resins studied by pmr.

Table 8. Proton Nmr Chemical Shifts of Methylene Groups in Phenolic Resins

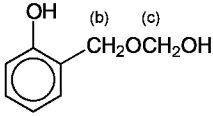
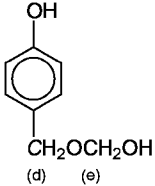
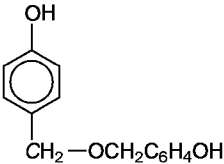
Methylene group	Chemical shift ² , ppm
2-CH ₂ OH	5.1
2-CH ₂ OR	5.0
4-CH ₂ OH	4.8
4-CH ₂ OR	4.7

2,2'-CH ₂	4.2
2,4'-CH ₂	4.1
4,4'-CH ₂	3.8
2-CH ₂ N	4.0
4-CH ₂ N	3.5

^a 10⁰ ◦ Concentration in *d*₅-pyridine.

A great wealth of microstructural information is provided by Fourier transform ¹³C-nmr. Using the much greater chemical-shift range of this technique, detailed structural information is provided for both the aliphatic and the aromatic carbons (Table 9). Current techniques provide highly reliable quantitative data and relative peak areas (18,53–57) and make possible a quantitative measure of the numbers of branch points and end groups. Branching can cause early gelation in a novolak resin, and end groups usually have greater reactivity in the thermosetting reaction than the backbone units. Another important advantage of ¹³-nmr is the parametric predictability of the chemical-shift values. As a result, unknown peaks can be assigned to a hypothetical structure with reasonable certainty. At the same time, the process can be reversed and a computer can provide detailed structural analysis. ¹³C-nmr has been applied to cured phenolic resins (55).

Table 9. Chemical Shifts of Methylene Carbons in Liquid Resoles

Structure ^a	Chemical shift ^b , ppm
methylol C in (8)	61.3 65.4 (b) 88.0 (c)
	
benzyl C in (9)	68.9
methylol C in (11)	63.8 68.5 (d) 88.0 (e)
	
	71.5
methylene C in (5)	31.5
methylene C in (6)	35.0
methylene C in (7)	40.4

^a Designated carbon is shown in italics or described.

^b From tetramethylsilane in *d*₆-acetone solution.

Chromatography. Gel-permeation chromatography (gpc) is an invaluable technique for determining the molecular size distribution of polymers. Phenolic resins, which have molecular weight components ranging from 100 to rarely more than 5000, require special column arrangements to optimize resolution. By using proper instrument calibration, it is possible to obtain number-average (*M_n*) and weight-average (*M_w*) molecular weight as well as quantitative information on free monomer and certain other low molecular weight species (58–60).

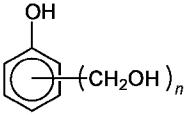
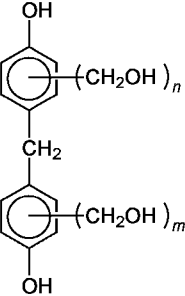
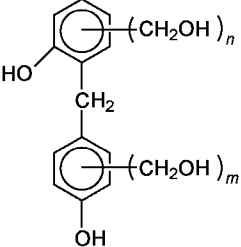
Many resole resins exist as phenolate salts in solution. Because these ionic species are sparingly soluble in carrier solvents such as tetrahydrofuran, careful neutralization and filtration are required. Although gpc is an excellent technique for examining medium and high molecular weight fractions, gc and high performance lc are more effective for analyzing low molecular weight species.

Gas chromatography (gc) has been used extensively to analyze phenolic resins for unreacted phenol monomer as well as certain two- and three-ring constituents in both novolak and resole resins (61). It is also used in monitoring the production processes of the monomers, eg, when phenol is alkylated with isobutylene to produce butylphenol. Usually, the phenolic hydroxyl must be derivatized before analysis to provide a more volatile compound. The gc analysis of complex systems, such as resoles, provides distinct resolution of over 20 one- and two-ring compounds having various degrees of methylation. In some cases, hemiformals may be detected if they have been properly capped (53).

The combined techniques of gas chromatography mass spectrometry (gc ms) are highly effective in identifying the composition of various gc peaks. The individual peaks enter a mass spectrometer in which they are analyzed for parent ion and fragmentation patterns, and the individual components of certain resoles are completely resolved.

High performance liquid chromatography (hplc) is extremely effective in separating individual resin components up to a molecular weight of 1000 according to size and polarity. Dilute-solution conditions and low temperatures preserve the structure of unstable components. The resins are usually not derivatized. Gradient solvent elution gives excellent peak separation (61,62). In one study, resoles catalyzed by sodium and barium hydroxide were compared, and the components were separated up to and including methylolated four-ring compounds (53). Resole components resolved by gc and hplc techniques are shown in Table 10. Like gc, hplc is most effective when combined with other analytical tools, such as mass and uv spectroscopy. By using preparative-scale hplc, individual peaks can be analyzed by proton and ¹³C-nmr.

Table 10. Resole Components Resolved by Hplc and Gc

Resin ^a	Components
	2; 4; 2,4; 2,6; 2,4,6
	2; 2,6; 2,2'; 2,6,2'; 2,6,2',6'
	2'; 6; 4; 2',6'; 6,2'; 4,2'; 4,6; 4,2',6'; 4,6,2'; 4,6,2',6'

^a Also, (5), (6), and (7), ie, 2,2'-, 2,4'-, and 4,4'-DPM.

Thermal Analysis. The main thermal analysis techniques applied to phenolic resins are tga and dsc. In tga, the sample weight is monitored microanalytically with time and temperature in air or nitrogen. When applied to resins and molding compounds, the scans indicate cure and decomposition temperatures accompanied by a measurable loss in weight. Resoles and novolaks lose from 5–20% of their weight on curing at 100–200°C. Weight loss provides information on shrinkage, void formation, and density change of composites.

Phenolic resins give a high char yield on combustion and tga provides a measure of the expected yield. Typical values are between 40 and 65% in nitrogen. Decomposition begins at 350°C and continues up to 600°C. Autoignition temperature in air is above 900°C. Thermogravimetric analyses have played an important part in the development of carbon–carbon and carbon–graphite-fiber composites containing phenolic resins. These composites are used in aircraft brake linings and carbon-pipe applications. In dsc and dta, heat flow and sample temperature are compared to a reference material. Glass-transition temperature is determined by dsc. The *T_g* of liquid resoles is below RT, that of friable novolaks is in the range of 50–75°C, and that of lightly cross-linked phenolics is between 150 and 225°C.

Cure kinetics of thermosets are usually determined by dsc (63,64). However, for phenolic resins, the information is limited to the early stages of the cure because of the volatiles associated with the process. For pressurized dsc cells, the upper limit on temperature is ca 170°C. Differential scanning calorimetry is also used to measure the kinetics and reaction enthalpies of liquid resins in coatings, adhesives, laminations, and foam. Software packages that interpret dsc scans in terms of the cure kinetics are supplied by instrument manufacturers.

Dynamic Mechanical Analysis. In dynamic mechanical analysis (dma), a vibrating or oscillating sample is heated at a programmed rate or held isothermally at elevated temperature. The frequency and damping characteristics of the sample are monitored with time. A change such as gelation or passage through the *T_g* causes abrupt changes in the fundamental oscillation frequency of the sample and the damping ability of the specimen. The oscillation frequency can be related to the storage modulus of the sample, whereas the damping contains information related to the loss modulus.

Softening and cure is examined with the help of a torsional pendulum modified with a braid (65), which supports thermosets such as phenolics and epoxies that change from a liquid to a solid on curing. Another method uses vibrating arms coupled to a scrim-supported sample to measure storage and loss moduli as a function of time and temperature. An isothermal analytical method for phenolic resins provides data regarding rate constants and activation energies and allows prediction of cure characteristics under conditions of commercial use (47).

A dma scan of a resole cure is shown in Figure 3. The sample is a glassy solid initially and the dma shows the distinctive *T_g* at 50°C followed by the appearance of a liquid state. Above 100°C, the liquid resin is vitrified by the chemical curing reactions, and gelation is observed at 160°C. Further reaction continues until 200°C, when a highly cross-linked, infusible solid is obtained. As the sample is cooled to RT, a slight increase in storage modulus is observed. The absence of a peak in the damping curve indicates that the *T_g* of the cross-linked systems exceeds the 225°C limit of the scan.

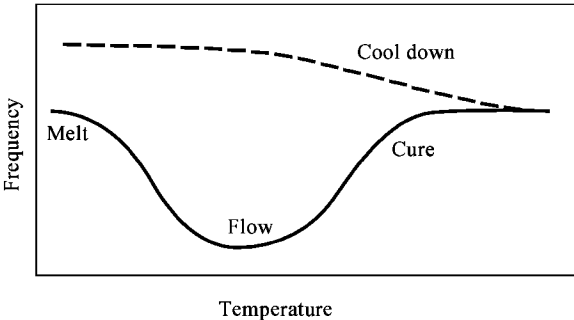


Fig. 3. Programmed dma scan of a resole phenolic resin; heating rate is 5°C min.

Dynamic mechanical analysis provides a useful technique to study the cure kinetics and high temperature mechanical properties of phenolic resins. The volatile components of the resin do not affect the scan or limit the temperature range of the experiment. However, uncured samples must be

supported by a braid, a scrim, or paper. This does not influence the kinetic results and can be corrected in the calculations of mechanical properties.

Control Tests. Numerous chemical and physical tests are used in the manufacture of phenolic resins to ensure correct properties of the finished resins, including the following: refractive index is used to estimate the dehydration during manufacture and is proportional to the solids content; viscosity is used to determine molecular weight and solids content; nonvolatiles content is roughly proportional to polymer content; miscibility with water depends on the extent of reaction in resoles; specific gravity is measured for liquid resins and varnishes; melting point of novolaks and solid resoles affects application performance; gel times determine the reactivity of the resins; resin flow is a measure of melt viscosity and molecular weight; particle size affects performance and efficiency; and flash point and autoignition temperature provide flammability-characteristic measurements required by government agencies regulating safety and shipping.

Health, Safety, and Environmental Factors

The factors contributing to the health and safety of phenolic resin manufacturing and use are those primarily related to phenol (qv) and formaldehyde (qv). Unreacted phenol in a resin can range from >10% for liquid resoles used in impregnation processes to <1% for novolaks intended for use as epoxy hardeners. Free formaldehyde can be 2–4° in liquid adhesives. Novolaks are usually free of formaldehyde. The toxicity of the resins is significantly lower than that of the phenol and formaldehyde starting materials. No detrimental toxicological effects have been reported for cured phenolic resins, which can be used in direct contact with food as in can coatings.

Uncured resins are skin sensitizers and contact should be avoided, as well as breathing the vapor, mist, or dust. Novolak-based pulverized products generally contain hexamethylenetetramine, which may cause rashes and dermatitis. Phenolic molding compounds and pulverized phenolic adhesives must be controlled as potentially explosive dusts. In addition, they contain irritating or toxic additives.

Phenol. Phenol monomer is highly toxic and absorption by the skin can cause severe blistering. Large quantities can cause paralysis of the central nervous system and death. Ingestion of minor amounts may damage kidneys, liver, and pancreas. Inhalation can cause headaches, dizziness, vomiting, and heart failure. The threshold limit value (TLV) for phenol is 5 ppm. The health and environmental risks of phenol and alkylated phenols, such as cresols and butylphenols, have been reviewed (66).

Formaldehyde. The toxicology and possible carcinogenicity of formaldehyde have been a matter of intense research. Formaldehyde is classified as a probable human carcinogen by the International Agency for Research on Cancer (IARC) and as a suspected human carcinogen by the American Conference of Governmental and Industrial Hygienists (ACGIH); the latter has lowered its TLV to 0.3 ppm. The Occupational Safety and Health Administration (OSHA) has set its time-weighted average for eight hours at 1.0 ppm and its short-term exposure level at 2.0 ppm (67).

Gaseous formaldehyde in concentrations above 1 ppm is extremely irritating to the mucous membranes. Aqueous formaldehyde is a protoplasmic poison and causes irritation of skin, eyes, nose, and throat. As a preservative it attacks bacteria and reacts with the amino groups of proteins. A 1985 study by the U.S. National Cancer Institute showed no evidence of human carcinogenicity. However, a thorough analysis of this study (68) showed a significantly elevated relative risk of lung cancer for workers exposed to formaldehyde concentrations higher than 0.5 ppm.

Wastewater. Phenol is a toxic pollutant to the waterways and has an acute toxicity (~5 m g L) to fish. Chlorination of water gives chlorophenols, which impart objectionable odor and taste at 0.01 mg L. Biochemical degradation is most frequently used to treat wastewater containing phenol. Primary activated sludge, along with secondary biological treatment, reduces phenol content to below 0.1 mg L (69).

Flammability. Phenolics have inherently low flammability and relatively low smoke generation. For this reason they are widely used in mass transit, tunnel-building, and mining. Fiber glass-reinforced phenolic composites are capable of attaining the 1990 U.S. Federal Aviation Administration (FAA) regulations for total heat release and peak heat release for aircraft interior facings (1,70).

Recycling. Thermosets are inherently more difficult to recycle than thermoplastics and thermosetting phenolics are no exception. However, research in this area has been reported, and molded parts have been pulverized and incorporated at 10–15° in new molding powders. Both German and Japanese groups had instituted this type of practice in 1992 (71,72) (see RECYCLING).

Economic Aspects

In 1993, worldwide consumption of phenolic resins exceeded 3 × 10⁶ t; slightly less than half of the total volume was produced in the United States (73). The largest-volume application is in plywood adhesives, an area that accounts for ca 49° of U.S. consumption (Table 11). During the early 1980s, the volume of this application more than doubled as mills converted from urea–formaldehyde (UF) to phenol–formaldehyde adhesives because of the release of formaldehyde from UF products. Other wood bonding applications account for another 15° of the volume. The next largest-volume application is insulation material at 12° .

Table 11. Long-Term Use of Phenolic Resins^a, 10³ t

End use market	1973	1983	1993
bonding			
abrasives	11	12	12
wood bonding	42	80	215
friction materials	15	12	22
foundry	50	26	39
insulation	112	165	166
<i>total</i>	<i>230</i>	<i>295</i>	<i>454</i>
laminating			
building	26	17	25
electrical	7	12	13
furniture	17	8	10
other laminating	3	30	30
<i>total</i>	<i>53</i>	<i>67</i>	<i>78</i>
plywood	125	580	687
molding compounds	173	100	74
protective coatings	10	10	6
export	14	8	14
others	49	55	79
<i>Total bonding and laminating</i>	<i>654</i>	<i>1115</i>	<i>1392</i>

^a Ref. 74.

As a mature industry, U.S. production and application of phenolic resins have paralleled the growth in the GNP. Only the consumption of phenolic resins for coatings and molding powders has decreased in recent years. The driving force behind coatings has been the need to reduce the volatile organics content (VOC) and the growth in alternative coatings such as waterborne, powder, and radiation-cured coatings. In the area of molding powders, numerous new engineering thermoplastics and alloys having superior properties have been introduced since the 1970s; however, phenolic molding powders continue to be used extensively in electrical, electronic, houseware, and appliance applications.

U.S. phenolic resin manufacturers include AlliedSignal Inc. Bendix; Ashland Chemical, Inc.; Borden, Inc.; Dexter Corp.; Dyno Polymers; Georgia-Pacific Corp.; Neste Resins Corp.; Occidental Chemical Corp.; Owens-Corning Corp.; Plastics Engineering Co.; PMC, Inc.; Resinoid Engineering

Co.; Spurdock Adhesives, Inc.; Stuart-Ironsidés, Inc.; and Valite Division of Valentine Sugars, Inc.

Prices of phenolic resins vary substantially depending on the application. In 1995, the price of general-purpose and semisolids was \$1.50–\$1.80 kg, whereas epoxy-hardener grades can exceed \$2.20 kg. Because raw materials of phenolic resins are derived from crude oil and natural gas, the prices of phenolic resins depend on the prices of these resources.

Applications

Coatings. For coatings applications, phenolic resins are grouped into four classes, depending on heat reactivity and the type of phenol. Substituted phenols are more compatible with oil and hydrocarbons, whereas heat-reactive resins require polar solvents. Depending on its nature, the resin can be used alone or as a modifying resin that acts as an adhesion promoter, a chemical cross-linker, or a hardening agent. In these cases the primary resin may be an alkyd, polyester, or epoxy (75).

Unsubstituted heat-reactive resins are designed for baked-on coatings and are usually not oil-soluble. They require strong solvents and, although most are not water-miscible, their low molecular weight, high formaldehyde content promotes water miscibility. These resins are available as solids, viscous liquids, and solutions. Resins prepared with an alkaline catalyst and a slight excess of formaldehyde over phenol are heat-reactive, but not as much as resole resins designed for fiber bonding and paper impregnation. Recommended cure conditions are 30 min and 150°C.

Heat-reactive resins are more compatible than oil-soluble resins with other polar-coating resins, such as amino, epoxy, and poly(vinyl butyral). They are used in interior-can and drum linings, metal primers, and pipe coatings. The coatings have excellent resistance to solvents, acids, and salts. They can be used over a wide range of temperatures, up to 370°C for short periods of dry heat, and continuously at 150°C. Strong alkalis should be avoided.

The maximum recommended film thickness is 25 µm. At greater thicknesses, volatiles from the curing reaction, mainly water and some formaldehyde and phenol, can cause defects. These coatings have excellent electrical insulation properties, ie, up to 20 V µm, because of low moisture absorption and low conductance. The coatings are hard with low flexibility, depending on curing conditions and film thickness.

Phenolic baking coatings can be used for metal, ceramic, and plastic surfaces. Applications include equipment for heating and air conditioning, chemical processing, petroleum refining, and water treatment. Some types are used in oilwell pipes and marine environments (76). Certain coatings can be used in food and beverage processing, subject to regulations.

Unsubstituted resins that are not heat-reactive have limited use in coatings because they do not form films. For other applications, they are cured with hexa to form the final thermoset bond. However, the low solubility of hexa in common coating solvents precludes its use as a hardener. The largest use of novolak resins in coatings is as a hardener for epoxy resins. The epoxy is frequently based on bisphenol A or epoxidized novolaks. Basic catalysts such as benzyldimethylamine are required for moderate baking conditions, such as 2 h at 180°C. The phenolic:epoxy ratio is adjusted to be stoichiometrically equivalent in order to give a highly cross-linked coating that has moderate flexibility and excellent resistance to chemicals, heat, and moisture. Powder coating of pipes continues to be a growing application, especially when corrosion resistance is required. In solution form, the epoxy–phenolic systems are used as metal primers and in pipe coating.

Substituted heat-reactive resins are most widely used in contact-adhesive applications and, to a lesser extent, in coatings (77,78); *p*-butylphenol, cresol, and nonylphenol are most frequently used. The alkyl group increases compatibility with oleoresinous varnishes and alkyds. In combination with these resins, phenolics reduce water sensitivity. Common applications include baked-on and electrical insulation varnishes, and as modifiers for baking alkyds, rosin, and ester gum systems. Substituted heat-reactive resins are not used for air-dry coatings because of their soft, tacky nature in the uncured state; substituted nonheat-reactive phenolics are the modifying resin of choice in this case.

Substituted nonheat-reactive resins do not form a film and are not reactive by themselves, but are excellent modifier resins for oleoresinous varnishes and alkyds. Their high glass-transition temperature and molecular weight provide initial hardness and reduce tack; oxygen-initiated cross-linking reactions take place with the unsaturated oils.

Early phenolic resin drying-oil varnishes were cooked in order to incorporate the phenolic resin into the formula. These resins have been replaced by cold-cut resins that reduce atmospheric emission by permitting direct incorporation of the phenolic after the oleoresinous varnish has been prepared. High solids systems enable coatings to meet the VOC standards required by regulatory agencies. Newer phenolic varnishes, developed in the 1980s, may contain as high as 80% nonvolatile solids (79).

Oleoresinous phenolic varnishes are excellent coatings that dry in 2–4 h and show exterior durability, corrosion resistance (especially when aluminum-modified), compatibility, solubility, and good package stability. Recoatability and intercoat adhesion are also excellent. The films are sensitive, however, to strong solvents and concentrated acids and alkalis. Unlike systems containing cross-linking phenolics, films containing these resins remain flexible.

Phenolics that are not heat-reactive may be incorporated into both air-dried and baked oleoresinous coatings. Applications vary widely and include clear and pigmented exterior varnishes, aluminum-maintenance paints, zinc-rich primers, can coatings, insulation varnishes, and concrete paints. As modifiers in a great variety of applications, they enhance the performance of oleoresinous and alkyd coatings.

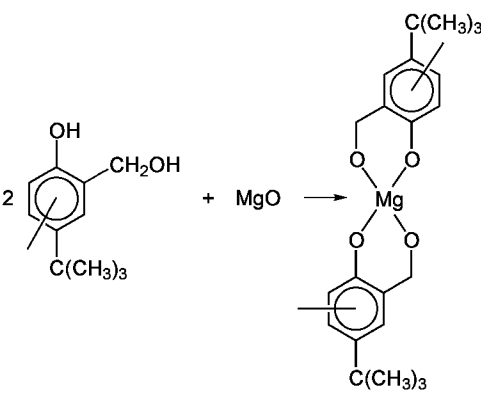
Dispersions. In phenolic resin dispersions, the continuous phase is water or a nonpolar hydrocarbon solvent. The resin exists as droplets that have particle sizes of 1–20 µm and are dispersed in the continuous phase. Aqueous dispersions are prepared either *in situ* during the preparation of the resin itself or by high shear mixing (25,35).

Aqueous dispersions are alternatives to solutions of liquid and solid resins. They are usually offered in 50% solids and may contain thickeners and cosolvents as stabilizers and to promote coalescence. Both heat-reactive (resole) and nonheat-reactive (novolak) systems exist that contain unsubstituted or substituted phenols or mixtures. A related technology produces large, stable particles that can be isolated as discrete particles (44). In aqueous dispersion, the resin structure is designed to produce a hydrophobic polymer, which is stabilized in water by an interfacial agent.

Aqueous dispersions are used in fiber bonding, paper coating, friction and abrasive applications, and laminates and wood bonding. Phenolic dispersions improve the strength of latex-contact adhesive applications. Epoxy-modified phenolic dispersions are prepared by dispersion of the phenolic epoxy resin. The systems are used for baked primer applications and bonding requirements. Minimum baking conditions are 20 min at 150°C (25).

Adhesives. Contact adhesives are blends of rubber, phenolic resin, and additives supplied in solvent or aqueous dispersion form; they are typically applied to both surfaces to be joined (80). Evaporation of the solvent leaves an adhesive film that forms a strong, peel-resistant bond. Contact adhesives are used widely in the furniture and construction industries and also in the automotive and footwear industries. The phenolic resins promote adhesion and act as tackifiers, usually at a concentration of 20–40%. In solvent-based contact adhesives, neoprene is preferred, whereas nitrile is used in specialty applications. The type and grade of phenolic resin selected control tack time, bond strength, and durability.

Neoprene–phenolic contact adhesives, known for their high green strength and peel values, contain a resole-type resin prepared from 4-*t*-butylphenol. The alkyl group increases compatibility and reduces cross-linking. This resin reacts or complexes with the metal oxide, eg, MgO, contained in the formulation, and increases the cohesive strength of the adhesive. In fact, the reactivity with MgO is frequently measured to determine the effectiveness of heat-reactive phenolics in the formulation.



Phenolic resin substantially increases open time and peel strength of the formulation (80). For example, higher methylol and methylene ether contents of the resin improves peel strength and elevated temperature resistance. Adhesive properties are also influenced by the molecular weight distribution of the phenolic; low molecular weight reduces adhesion (82).

Waterborne contact adhesives contain an elastomer in latex form, usually an acrylic or neoprene-based latex, and a heat-reactive, cross-linkable phenolic resin in the form of an aqueous dispersion. The phenolic resin improves metal adhesion, green strength, and peel strength at elevated temperature. A typical formulation contains three parts latex and one part phenolic dispersion (dry weight bases). Although metal oxides may be added, reaction of the oxide with the phenolic resin does not occur readily.

Bonding properties of water-based contact adhesives are similar to those of solvent-based systems, but are free of flammability hazards. However, drying times are longer and the bond is sensitive to moisture.

Carbonless Copy Paper. In carbonless copy paper, also referred to as pressure-sensitive record sheet, an acid-sensitive dye precursor, such as crystal violet lactone or *N*-benzoylleucomethylene blue, is microencapsulated with a high boiling solvent or oil within a cross-linked gelatin (76,83,84) or in synthetic mononuclear microcapsules. Microcapsules that have a starch binder are coated onto the back of the top sheet. This is referred to as a coated-back (CB) sheet. The sheet intended to receive the image is treated on the front (coated-front (CF)) with an acid. When the top sheet is mechanically impacted, the dye capsules rupture and the dye solution is transferred to the receiving sheet where the acid developer activates the dye.

The original acid–clay developers have been largely replaced by phenolic compounds, such as para-substituted phenolic novolaks. The alkyl group on the phenolic ring is typically butyl, octyl, nonyl, or phenyl. The acidity is higher than that of a typical unsubstituted novolak because of the high concentration of 2,2'-methylene bridges.

Color intensity and permanence are improved by metal carboxylate salts, especially zinc salts (83), which catalyze the dye development and stabilize the dye in its colored form. The substituted novolak resin, along with extender and binder, can be applied to the receiving sheet as a solution or aqueous dispersion. Aqueous dispersions are probably the most widely used; they are manufactured by the resin supplier or the user from the base resin.

A typical coating composition for the CF component is shown in Table 12. It is dried in a high velocity air oven at 93°C.

Table 12. CF Coating Slurry Formulation

Constituent	Parts
kaolin clay	64
CaCO ₃	3
colloidal silica	5.4
hydroxyethyl starch	3
styrene–butadiene latex	12
novolak resin dispersion	12

Molding Compounds. Molding compounds were among the earliest applications for solid phenolic resins. Molding neat phenolic resin was almost impossible and the strength, especially on impact, was poor without reinforcement. Combining the resin, usually a novolak, with hexa, wood-flour, or asbestos reinforcement, as well as pigments and additives, gave a moldable thermoset. The molded articles exhibit high temperature, flame, and chemical resistance, retention of modulus at elevated temperature, and hardness. Systems that have good electrical properties can be formulated at low cost. Resins can be compounded with a choice of fillers for a variety of applications.

The requirements for 25 molded compounds are given in ASTM D700-81, ranging from general-purpose to impact-, heat-, and flame-resistant grades. The general-purpose compounds, having good strength and electrical properties, are used in electrical appliances, automotive ignitions, and handles. Impact-resistant grades contain longer glass and textile fibers and rubber modifiers. Applications include pump housing and switches. Heat resistance is improved by incorporating mineral and glass fillers. Electrical-grade compounds contain mica and fiber glass fillers, and are used in specialty switches and cases. Graphite filler reduces the coefficient of thermal expansion. Although phenolic molding compounds compete with a wide variety of thermoset and thermoplastic compounds, they continue to be the preferred material in numerous applications, such as reinforced phenolic pistons in automotive disk brakes and transmission torque converters. Table 13 lists several applications for molded phenolic parts.

Table 13. Applications for Molded Phenolic Parts^a

Market	Key features	Examples
appliances	heat, moisture, and chemical resistance; thermal and electrical insulation	coffee pot handle and base; steam iron parts; dishwasher components
automotive	heat, moisture, and hydro-carbon resistance; dimensional stability	brake caliper pistons; trans-mission parts; belt pulleys
electrical	heat and wear resistance; thermal and electrical insulation	circuit breakers; motor control parts; solenoid covers
aerospace	heat and chemical resistance; rigidity	nozzle inserts; ablative surfaces
packaging	dimensional stability; chemical resistance	closures

^a Ref. 85.

Phenolic molding materials are prepared by first blending the pulverized raw materials. In a compounding operation, the blend is passed through heated rolls to form a sheet as the resin melts. The compounded mixture remains on heated rolls until the desired flow and viscosity are reached. At this point, it is discharged as a thin sheet and allowed to cool. The product is then granulated and packaged. Alternatively, a Banbury-type mixer can be used for the initial mixing; the mix is then fed to a roll mill for finishing and sheeting. Extrusion processes are continuous and provide a pelletized product (see

MIXING AND BLENDING).

Phenolic resins used in molding materials are predominantly novolaks with hexa as the curing agent. Oxalic acid-catalyzed novolaks and high ortho resins are used. Resole resins may be used in place of hexa as the curing agent in certain electrical applications in which the NH₃ generated from the hexa can adversely affect metals such as copper and silver. Typically, the compound contains 40–50 wt % resin binder, 40–45% filler, and 5–10% pigment and additive. Fillers and reinforcements include wood flour, nutshell flour, cellulose fiber, mica (qv), wollastonite, mineral wool, mineral flour, glass fiber, organic fiber, carbon fiber, clay, and talc (qv). Additives and pigments are magnesium oxide, graphite, molybdenum sulfide, stearates, calcium carbonate, carbon black, nigrosine, lime, fluoropolymers, and salicylic acid.

Molding materials are fabricated into articles by compression, transfer, and injection molding processes. For compression molding, the powder is poured into the mold cavity or a preform heated prior to molding. Temperature ranges from 140–190°C and pressures from 13.8–55.5 MPa (2000–8000 psi). Because the material is more fluid in the mold, transfer molding is suitable for molding around inserts or for intricate shapes. In injection molding, cross-linking in the barrel and sprues must be avoided.

Abrasives. Abrasive materials are either bonded or coated. Bonded phenolic abrasives have superior strength and shock resistance compared to sintered ceramic compositions. The higher stability permits higher rotational speeds for resin-binder wheels; however, temperatures are lower than with ceramic wheels.

Synthetic aluminum oxide and silicon carbide are the principal inorganic abrasive materials. They are used in grit sizes from 44 or finer to 1680 μm (12–325 mesh). The coarser grit sizes are used for rough work when wheels are as large as 60 cm in diameter and 7.5-cm thick. These wheels are mounted in swinging-frame grinders and used to remove surface imperfections from stainless steel billets. Finer grit (250 μm (60 mesh)) may be used in 45-cm-dia-thick cutoff wheels mounted in automatic grinders. These wheels can slice through a solid 2.5-cm steel bar in 4 s, leaving a clean, shiny cut surface.

In the usual method of manufacture, the grit is first wetted with a low viscosity liquid resole phenolic resin or furfural, using 1–3 wt % of the grit. A dry mixture of phenolic resin and fillers is added and the combination of wet grit and powder is tumbled until each grit particle is coated. The mix is temporarily dry and free-flowing. The powdered phenolic resin varies from 6 to 10% of the weight of the grit. The quantity and type of filler (alumina, cryolite, iron oxide, silicate) depend on the intended use of the wheel, whose grinding characteristics are affected by the filler. The free-flowing grit resin mix is placed into the mold and pressed at RT for 1–2 min at 13.8–34.5 MPa (2000–5000 psi). The dense form is removed and cured for 12–24 h; temperatures are increased gradually to 185°C for 8–12 h. Pressing temperatures may be increased for high density structures.

Heat resistance is an important characteristic of the bond. The strength of typical abrasive structures is tested at RT and at 300°C. Flexural strengths are between 24.1 and 34.4 MPa (3500–5000 psi). An unmodified phenolic resin bond loses about one-third of its room temperature strength at 298°C. Novolak phenolic resins are used almost exclusively because these offer heat resistance and because the moisture given off during the cure of resole resins results in undesirable porosity. Some novolaks modified with epoxy or poly(vinyl butyral) resin are used for softer grinding action.

Coated abrasives, such as sheets, disks, and drums, are used for polishing and finishing. Here, too, the abrasives, such as aluminum oxide and silicon carbide, have replaced the flint and garnet of common sandpaper. These industrial coated abrasives are manufactured from cloth or tough paper base. First, a coat of medium viscosity liquid resole resin is laid down on a continuous web of the backing. The web passes wet-side-down over a pan of grit, which adheres to it electrostatically and remains embedded in the resin layer. Altering the strength of the applied electrostatic field and the speed of the web can control the amount of grit deposited, which varies from an open to a closed dense mass of abrasive. The uncured coated sheet is partially dried in a low temperature oven at 60°C, and a second, thinner coating of lower viscosity liquid resole resin is applied as a top to anchor the grit thoroughly. The coated web is taken off in rolls to be cured at 107–120°C for 3–4 h.

The resins should dry quickly and cure well at low temperatures. They usually are made at a high pH with high ratios of formaldehyde to phenol and held to fairly low molecular weight. Typical viscosities are 15,000 mPas(cP) at 75% solids content for a first coat and 1000 mPas(cP) at 50% solids for the top resin. For dense backing materials, such as fiber disks, a typical resin has a viscosity of 50,000 mPas(cP) at 80% solids and is cured at 148°C. New opportunities are arising for coated abrasive belts in high pressure grinding. High strength backing has been developed and zirconia media are available for grinding.

Friction Materials. Phenolic friction materials are made from molding compounds developed to meet the extraordinary demands required by friction elements in the transportation industries. Friction materials are used for brake linings, clutch facings, and transmission bands. A moderately high coefficient of friction, which is temperature-independent, is needed. In addition, the material must be high in strength, low in wear and abrasion, and resistant to moisture and hydraulic fluids.

In the 1980s, significant changes occurred in automotive brake elements. Reformation of friction elements eliminated asbestos and increased employment of disk brakes. These developments required binder resins that have higher temperature performance without affecting the coefficient of friction and wear. Alternatives to asbestos include mineral, carbon, aramid, and metal fibers, especially iron. Semimetallic linings contain as much as 70% iron fibers. Other inorganic fillers include barites, alumina, lime, magnesia, and clay. Graphite and molybdenum disulfide are used to reduce scoring.

The resins can be a novolak–hexa or a resole–novolak blend. In some applications liquid resoless are used. Addition of alkylated phenol, oil, or cashew nutshell liquid (CNSL) reduces hardness and increases abrasion resistance. Modification by rubber improves the coefficient of friction and reduces brake fading.

Many friction material formulations contain 5–15 wt % of friction particles, the granulated cross-linked products of the reaction of CNSL, a phenol substituted at the meta position with a C₁₅, unsaturated side chain, and formaldehyde. Friction particles range in size from 50 to 500 μm. They reduce frictional wear and increase pedal softness (86).

Manufacture of friction elements includes the impregnation of fabrics and subsequent lamination, the wet-dough process, and the dry-mix process. Elements from the last two are prepared by compression-molding the formulation for up to an hour at 150–175°C. Thick brake elements require a carefully controlled heating-and-cooling cycle to minimize stresses created by expansion and contraction (see BRAKE LININGS AND CLUTCH FACINGS; FILLERS).

Foundry Resins. In the foundry industry, phenolic resins are used as the binder for sand in the manufacture of shell molds and cores. The two mating halves are joined by clamps or a bonding agent to form a shell mold into which the molten metal is poured for castings. The shell is formed by depositing a resin–sand mix on a hot metal pattern plate. After a certain period the pattern is inverted and the excess resin sand is removed. The sand particles are bonded by an oven cure, and the shell is ejected from the pattern plate.

Iron is the preferred metal for casting; steel and nonferrous metals are used in smaller amounts. Most castings are made in green sand molds, ie, uncured molds of sand, clay, and water. However, the use of shell moldings is growing, because such moldings permit reproducibility of castings with close dimensional accuracy. In addition, the simplicity of equipment procedures reduces costs.

The shell-molding process, introduced in the United States in 1948, is an important market for phenolic resins. In the original process, dry sand and powdered resin (6–8%) are blended. However, because of the high binder content and the difficulty in obtaining a uniform mix, precoating methods were developed.

Both cold- and warm-coating processes employ solutions of phenolic resins. The principal process used for foundry resins is the hot-coating process. It is the fastest, least expensive, and safest process, and it requires no volatile removal. The sand is heated to 135–170°C in a muller, and solid novolak resin in flake form is added, which melts quickly and coats the sand. A lubricant may be added at this point. After one minute of mulling, the batch is cooled by adding water, which evaporates rapidly.

Hexa, which is not supplied with the resin, is usually added either with the water as a solution or just before or immediately after the water addition. By quenching the mix with water, the resin-coated sand is cooled to a point where there is no significant reaction with the curing agent. Any reaction between the resin and the hexa in the muller affects the bonding properties of the coated sand. As the batch cools and begins to break up, more lubricant may be added, which remains on the outside of the coated grains where it is most effective.

The total cycle time for most production batches is 2.5–3.5 min, considerably shorter than the cold- or warm-coating processes. Although a few

available units have large capacity per batch, high production rates are possible because of the short cycle. While one batch is being mulled, sand for the next batch can be heated. The typical flake phenolic is an intermediate melting-point novolak containing 4% lubricant.

Laminates. Laminate manufacture involves the impregnation of a web with a liquid phenolic resin in a dip-coating operation. Solvent type, resin concentration, and viscosity determine the degree of fiber penetration. The treated web is dried in an oven and the resin cures, sometimes to the B-stage (semicured). Final resin content is between 30 and 70%. The dry sheet is cut and stacked, ready for lamination. In the curing step, multilayers of laminate are stacked or laid up in a press and cured at 150–175°C for several hours. The resins are generally low molecular weight resoles, which have been neutralized with the salt removed. Common carrier solvents for the varnish include acetone, alcohol, and toluene. Alkylated phenols such as cresols improve flexibility and moisture resistance in the fused products.

Industrial laminates are composed of two or more webs in sheet form that have been impregnated with a thermosetting resin and molded under heat and pressure. The web is usually made of cellulose paper, cotton fabric, glass cloth, glass mat, asbestos paper, wood veneer, and similar materials. Laminates are manufactured in sheets, rods, tubes, and special shapes, which can be machined by common methods to produce parts for a wide range of uses. An important example is in printed-circuit applications, which require varnishes that impart electrical properties, punchability, machinability, and hot-solder and solvent resistance to the laminate. General-purpose laminates are used in gears, spacers, cams, ball-bearing retainer rings, and other structural parts requiring machinability as well as moisture, oil, and impact resistance.

A decorative laminate is composed of layers of resin-impregnated paper, which have been molded together under heat and pressure to form a solid homogeneous sheet. The component parts are composed of corestock impregnated with phenolic, print, and overday. The corestock confers body and strength. The print sheet provides the design characteristic; it is impregnated with melamine resin, as is the overday sheet that imparts abrasion protection to the print sheet. Approximately seven sheets of phenolic-impregnated corestock are covered with one print sheet, which, in turn, is covered with one overday sheet. This combination is molded against polished steel plates to obtain the desired finish, usually satin or gloss.

The type of varnish used in the process depends on the kraft paper manufacturer and basis weight of the papers; the machine, temperature, and control (scraper bars, squeeze rolls) used; the method of cutting the paper to size; the laminate being produced (post-forming or regular); and the press-cure cycle (see LAMINATED MATERIALS, PLASTIC).

Air and Oil Filters. Liquid resole resins are used to coat and penetrate the cellulose fibers of filters and separators in order to increase strength and stiffness and protect against attack by the environment. The type of phenolic to be used depends on both the final property requirements and the papermaking process.

Air and oil filters are made by a dry-web process in which the filter paper is dried over heated metal drums. The paper is saturated with the phenolic resin solution, either off- or on-line, and dried in an oven advancing the resin to the B-stage (semicured). The sheet, containing 20–30% resin, is rolled and shipped to the filter-unit manufacturer, where the sheet is convoluted and the filter assembled and cured to the C-stage (fully cured).

The resins used in air and oil filters are moderate-to-low molecular weight, catalyzed by caustic in one step; 10–20% alcohol is added; solids content is in the range of 50–60%. These resins are designed to penetrate the sheet thoroughly, yet not to affect the porosity of the paper. In the B-stage, the resin must have sufficient flexibility to permit pleating; the C-stage should have stiffness and resistance to hot oil.

Wood Bonding. This application requires large volumes of phenolic resins (5–25% by weight) for plywood, particle board, waferboard, and fiberboard. Initially, phenolic resins were used mainly for exterior applications, whereas urea–formaldehyde (UF) was used for interiors. However, the concern over formaldehyde emission has caused the replacement of UF by phenol–formaldehyde adhesives.

Different phenolic resins are used for different types of wood; for example, plywood adhesives contain alkaline-catalyzed liquid resole resins. Extension with a filler reduces cost, minimizes absorption, and increases bond strength. These resins have an alkaline content of 5–7% and are low in free phenol and formaldehyde. Because many resins have a high water content and limited storage stability, they are frequently made at or near the mill producing the plywood product. The plywood veneers are dried, coated with resin, stacked for pressing, and cured at 140–150°C.

Particle board and wood chip products have evolved from efforts to make profitable use of the large volumes of sawdust generated annually. These products are used for floor underlayment and decorative laminates. Most particle board had been produced with urea–formaldehyde adhesive for interior use; resin demand per board is high due to the high surface area requiring bonding. Nevertheless, substantial quantities of phenol–formaldehyde-bonded particle board are produced for water-resistant and low formaldehyde applications.

The phenolic resins used for particle board are NaOH-catalyzed resoles of low viscosity and high water miscibility, similar to the liquid resole adhesives used in plywood manufacture. The higher resin and caustic content of the board frequently necessitates the addition of hydrophobic agents such as wax emulsions to increase the barrier properties of the board. The adhesive is applied to the particles in thin streams using high agitation to maximize material usage. Boards are cured in presses for 5–10 min at 150–185°C.

Waferboard, a more recent wood construction product, competes more with plywood than particle board. Waferboard and strand board are bonded with solid, rather than liquid, phenolic resins. Both pulverized and spray-dried, rapid-curing resins have been successfully applied. Wafers are dried, dusted with powdered resin and wax, and formed on a caul plate. A top caul plate is added and the wafers are bonded in a press at ca 180°C for 5–10 min. Physical properties such as flexural strength, modulus, and internal bond are similar to those of a plywood of equivalent thickness.

Fiberboard or hardboard is made of low grade wood and wood waste. In the wet production process, a sheet is produced on a papermaking machine, such as a fourdrinier. A liquid resole is usually added to the beater section and precipitated onto the wood fibers by adjusting the pH. The moderately dry felt is further dried and cured in an off-line press.

Fiber Bonding. In fiber bonding, the resin is used as a binder in such products as thermal insulation batting, acoustic padding, and cushioning materials. All these materials consist of long fibers (glass or mineral fiber, cotton, polyester) laid down in a randomly oriented, loosely packed array to form a mat. They are bonded with resin to preserve the special insulating or cushioning quality of the mat.

In the dry-bonded process, fibers are reclaimed from woven textile scraps and powdered resin. Automotive acoustical padding (cotton and synthetic organic fiber) and thermal insulation batting (glass fiber) are made by this process. Inorganic and or organic fibers are picked open to allow distribution of the phenolic resin among individual fibers. The fiber-resin mixture is formed into a blanket by cross-lapping a series of webs or by being blown into a chamber for deposition. The blanket is oven-cured to fuse the resin. Oven temperatures, dwell times, molding temperatures, and cycle times vary, depending on the type of fiber and curing characteristics of the resin.

The wet-bonded process uses virgin fiber spun from molten glass and liquid water miscible resin. High grade, thermal insulation batting is made by this process. The most common wet-bonded process for manufacturing glass or rock-wool, thermal insulation batting employs steam- or air-blowing. A stream of molten glass falls from the melting furnace into a rapidly spinning platinum cylinder. Centrifugal force causes the glass to be extruded through many small holes in the cylinder wall. At the outside, the fibers are attenuated (drawn) by the action of a jet flame from burners located around the spinning cylinder. Spray nozzles located just below this unit apply liquid binder resin to the fibers, which are sucked onto a moving screen to form a continuous mat. The mat is conveyed through a forced hot-air curing oven similar to that used in the dry-bonded process. Curing is followed by cool-down, trimming, slitting, and roll-up.

Fiber bonding resins are marketed in the form of finely pulverized powders of novolak hexa mixtures or resoles and water-miscible liquids. Resole resins are the standard for the fiber bonding industry. They have excellent curing properties and low content of solvent extractables after curing. Bond strength to the fiber is high and resin usage is low. Solid resole resins must be stored cool to reduce sintering and aging.

Novolak-based resins are primarily used for organic fibers. They require a cross-linking reagent, usually hexa, for cure. Bond strength is good, producing bonded mats that have high tensile strengths. The storage stability is also good, so that refrigeration is rarely required. A typical binder mix produced by the customer consists of phenolic resin, urea, hydrocarbon oil or wax emulsion, ammonium sulfate, and occasionally a silane adhesion promoter; it is diluted with water to 5–10% solids content. The urea acts as a scavenger for the free formaldehyde (5–8%) invariably present, both as an inexpensive extender and, owing to its high nitrogen content, as an antimoldering agent. The oil or wax emulsion acts as a fiber lubricant and dust suppressant, the ammonium sulfate as a cure accelerator, and the silane as an adhesion promoter. The large volume of water in the binder mix cools the newly spun fibers, thus preventing premature resin cure, ie, before the mat has been formed and compressed to the proper thickness. All water-miscible

resins must be stored under refrigeration.

Composites. The history of phenolic resin composites goes back to the early development of phenolic materials, when wood flour, minerals, and colorants were combined with phenolic resins to produce molding compounds. In later applications, resin varnishes were developed for kraft paper and textile fabrics to make decorative and industrial laminates. Although phenolics have been well characterized in glass-reinforced composites, new developments continue in this area, such as new systems for liquid-injection molding (LIM) and sheet-molding compounds (SMC). More complicated composite systems are based on aramid and graphite fibers.

Epoxy resins are usually superior to phenolics because of better adhesion, lower shrinkage on curing, and a much lower volatile content. Nevertheless, there are many specific applications for phenolic resins, including formable laminates and sheet-molding compounds having low flammability and smoke generation. These systems are suitable for aircraft interiors. Another area is in binders for carbon–carbon composites for applications such as in ablative coatings of reentry vehicle and aircraft brakes. The high char yield and the strength and porosity of the char are important in these applications. Finally, phenolic novolaks are effective curing agents for epoxies, and epoxidized novolaks are being used in applications requiring high functionality and high cross-link density.

Glass-Reinforced Composites. These composites are prepared from SMC and by LIM. Phenolic resins are usually not amenable to LIM because of the volatiles generated. However, a resin suitable to LIM has been prepared (87). The properties of the 60% glass-mat-reinforced product obtained from this resin were compared to conventional isophthalic polyester and vinyl ester sheet-molding compounds at equivalent glass loading; the mechanical properties are shown in Table 14 (88). The phenolic resin systems have slightly lower strength values but measurably higher impact resistance. In addition, the phenolic glass composite has a UL-94 V0 rating, whereas the polyester SMC burns (see POLYESTERS, UNSATURATED).

Table 14. Mechanical Properties of Composites^a

Matrix resin	Flexural strength ^b , MPa ^c	Elongation, %	Notched Izod, J m ^d	UL-94
phenolic	228	2	1868	V0
polyester	276	1.75	960	burns
vinyl ester	310	2	1227	burns

^a Glass content, 60%.

^b For all three matrix resins, the flexural modulus is 12 GPa (1.7 × 10⁶ psi).

^c To convert MPa to psi, multiply by 145.

^d To convert J m to ftlb/in., divide by 53.38.

A patent describes the glass-reinforced pultrusion of phenolic resin composites (89). In pultrusion, bundles of continuous fiber are wetted with resin and shaped and cured through a series of heated dies. A sulfonic acid accelerates the cure rate, which must be high. The mechanical properties are reported to be equivalent to those obtained from conventional unsaturated polyesters, but the heat-distortion temperature and flammability resistance are superior. An acid-free phenolic pultrusion system has been announced that can produce large panels for buildings, ships, aircraft, and mine ducts. The system involves an injection-type die rather than conventional dip tanks, and up to 70% glass loadings can be attained (90).

Phenolic sheet-molding compound is seeing increased use in interior and exterior applications in mass transit. The trains in service on the English Channel tunnel have several panels made from phenolic SMC which had to meet the strict European safety standards for fire and smoke generation.

Phenolic resin prepregs on glass fibers are prepared from alcohol solution and occasionally from aqueous solutions. The glass web is dipped in a tank containing resin solution; the resin content in the web is controlled by metering rolls. The wet web is dried in a horizontal or vertical oven arrangement and cured to B-stage of the resin. The dry web is cut, stacked, and stored under refrigeration until used in a laminate. Depending on the application, the prepreg can be rigid, similar to the epoxy-type prepregs used in laminated circuit boards, or soft and flexible, which provides the tack and drape required to conform to a shaped article. Autoclave-molding is essential for shaped articles to reduce void formation by trapped air and volatiles. Examples of shaped articles include interior parts of commercial airplanes such as the composite duct assembly, based on wrapped phenolic–fiber glass, being used on the Boeing 737. Although not equal to epoxy-prepreg systems in strength, phenolic resins provide superior flammability resistance and lower smoke toxicity, because antimony and halogens are absent.

Carbon-Fiber Composites. Cured laminates of phenolic resins and carbon-fiber reinforcement provide superior flammability resistance and thermal resistance compared to unsaturated polyester and epoxy. Table 15 shows the dependence of flexural strength and modulus on phenolic–carbon-fiber composites at 30–40% phenolic resin (91). These composites also exhibit long-term elevated temperature stability up to 230°C.

Table 15. Strength Properties of Phenolic–Carbon-Fiber Composites

Property	Resin, %		Epoxy novolak, 27
	Phenolic	35	
tensile strength, MPa ^a	115	63	64
flexural strength, MPa ^a	183	126	110
flexural modulus, GPa ^b	15.8	6.3	6.4

^a To convert MPa to psi, multiply by 145.

^b To convert GPa to psi, multiply by 145,000.

Carbon–Carbon Composites. Above 300°C, even such polymers as phenolics and polyimides are not stable as binders for carbon-fiber composites. Carbon–carbon composites are used at elevated temperatures and are prepared by impregnating the fibers with pitch or synthetic resin, followed by carbonization, further impregnation, and pyrolysis (91).

Carbon–carbon composites are used in high temperature service for aerospace and aircraft applications as well as for corrosion-resistant industrial pipes and housings. Applications include rocket nozzles and cases, aircraft brakes, and satellite structures. Carbonized phenolic resin with graphite fiber functioned effectively as the ablative shield in orbital re-entry vehicles for many years (92).

Phenolic and furfuryl alcohol resins have a high char strength and penetrate into the fibrous core of the fiber structure. The phenolic resins are low viscosity resoles; some have been neutralized and have the salt removed. An autoclave is used to apply the vacuum and pressure required for good impregnation and sufficient heat for a resin cure, eg, at 180°C. The slow pyrolysis of the part follows; temperatures of 730–1000°C are recommended for the best properties. On occasion, temperatures up to 1260°C are used and constant weight is possible even up to 2760°C (93).

Liquid-Injection Molding. In liquid-injection molding (LIM), monomers and oligomers are injected into a mold cavity where a rapid polymerization takes place to produce a thermoset article. Advantages of these processes are low cost, low pressure requirement, and flexibility in mold configuration. Conventional systems, such as isocyanate with polyol, release little or no volatiles. The generation of substantial volatiles in the mold is obviously undesirable and has represented a significant obstacle to the development of a phenolic-based LIM system. A phenolic LIM system based on an

anhydrous high ortho liquid resole has been reported (94,95). Formaldehyde is present in the form of both methylol and phenolic hemiformals. A formaldehyde:phenol ratio of ca 1.5–1.0 is used at a pH of 4–7. Divalent salts catalyze the reaction at 80–90°C for 5–8 h and the water is removed by an azeotropic solvent. Uncatalyzed resoles have excellent storage stability.

An important aspect of this procedure is the use of latent acid catalysts, such as phenyl hydrogen maleate, phenyl trifluoroacetate, and butadiene sulfone. These catalysts reduce the peak exotherm from over 200°C to 130–160°C. The resin catalyst mixture has a working life of up to several days at RT. The elevated temperature of molding these latent catalysts generates the corresponding acids, namely, maleic, trifluoroacetic, and phenolsulfonic, which catalyze the resole reaction. Typically, a cycle time of 1–2 min is required for a mold temperature of ~150°C.

The water liberated during the cure has no apparent effect on the composite properties. Glass-filled composites prepared in this manner retain mechanical properties at elevated temperatures as well as solvent and flammability resistance (88). Phenolic–graphite-fiber composites that exhibit superior mechanical properties have also been prepared by this process.

Foam. Phenolic resin foam is a cured system composed of open and closed cells with an overall density of 16–800 g cm³. Principal applications are in the areas of insulation and sponge-like floral foam. The resins are aqueous resoles catalyzed by NaOH at a formaldehyde:phenol ratio of ca 2:1. Free phenol and formaldehyde content should be low, although urea may be used as a formaldehyde scavenger.

The foam is prepared by adjusting the water content of the resin and adding a surfactant (eg, an ethoxylated nonionic), a blowing agent (eg, pentane, methylene chloride, or chlorofluorocarbon), and a catalyst (eg, toluenesulfonic acid or phenolsulfonic acid). The sulfonic acid catalyzes the reaction, while the exotherm causes the blowing agent, emulsified in the resin, to evaporate and expand the foam (96). The surfactant controls the cell size as well as the ratio of open-to-closed cell units. Both batch and continuous processes are employed. In the continuous process, the machinery is similar to that used for continuous polyurethane foam.

The properties of the foam depend mainly on density and the cell character. For insulation, a high content of closed cells, along with an encapsulated fluorocarbon blowing agent, is desired. The foam must have sufficient strength and be able to resist smoldering, ie, a glowing combustion characteristic of early foams. Smoldering is controlled by additives and foam structure. For floral foam, lower density and an open-cell structure having low toughness are desirable. Floral foam usually contains a wetting agent in the formulation to ensure the rapid uptake of water.

When smoldering is eliminated, phenolic foams exhibit excellent flammability resistance. Their limiting oxygen index (LOI) is 32–36%, and their smoke density is lower than that of polyisocyanurate foam of similar density. High thermal insulation K values have been obtained on foams of small cell size containing fluorocarbon blowing agents. Phenolic foams are used for building products and warehouse insulation (see FOAMS; INSULATION).

Spheres. Hollow spherical fillers have become extremely useful for the plastics industry and others. A wide range of hollow spherical fillers are currently available, including inorganic hollow spheres made from glass, carbon, fly ash, alumina, and zirconia; and organic hollow spheres made from epoxy, polystyrene, urea–formaldehyde, and phenol–formaldehyde. Although phenol–formaldehyde hollow spheres are not the largest-volume product, they serve in some important applications and show potential for future use.

In an early process by Sohio, a variety of compositions were produced by dissolving a film-forming polymer in a solvent (or water), adding a blowing agent, and spray-drying the resulting solution under carefully controlled conditions. The use of spray drying to produce discrete, uniform hollow spheres, organic or inorganic, requires high skills in formulation, process control, and engineering. Some of the organic materials required include phenolic resins, poly(vinyl alcohol), polystyrene, methylcellulose, and protein. In the case of the phenolic resins, water-miscible resoles were used; water was the solvent and ammonium carbonate–ammonium nitrate or dinitrosopentamethylenetetramine was the blowing agent. Hollow phenolic spheres in sizes of 5–50 μm dia were obtained. The principal intended use was as a floating cap on petroleum naphtha and crude oil to retard evaporation.

Emerson & Cumming, Inc. eventually bought the rights to the Sohio process and produced a variety of microspheres. Union Carbide was licensed to produce the phenolic microspheres offered under the name Phenolic Microballoons (Table 16). When Phenolic Microballoons are introduced into a crude-oil storage tank, they form a fluid seal that rises and falls with the level of the oil. A continuous vapor-barrier seal is formed, which reduces evaporational losses up to 90%. Tests have been conducted under various mechanical and weather conditions and with crude oils of varying vapor pressure.

Table 16. Physical Properties of Phenolic Microballoons

Property	Value
average particle size, μm	43
typical particle range, μm	20–120
liquid displacement density, g cm ³	0.15–0.35
bulk density, g cm ³	0.07–0.15
toluene flotation, %	~90

Phenolic Microballoons applications in plastics take advantage of low density, porosity, and surface-to-volume ratio to produce lightweight parts. Probably the most notable example is the syntactic foam.

Whereas a conventional foam uses a blowing agent to form a cellular structure as the resin sets, syntactic foams are made by incorporating hollow spheres into liquid resins, especially epoxy resins in the case of Phenolic Microballoons, although urethanes and polyesters are also used. Having a continuous polymer phase and taking advantage of the high compressive strength of the spheres, syntactic foams can be made much stronger than conventional foams. In addition, they are being formed in place. As a result, syntactic foams have become widely used as a modeling, styling, and structural core material in automotive, marine, and aerospace applications.

A mixture of Phenolic Microballoons and resin binder has a putty-like consistency. It can be molded to shape, trowelled onto surfaces, or pressed into a core. Curing gives a high strength, low density (0.144 g cm³) foam free of voids and dense areas, and without a brittle skin. Syntactic foams are used in widely diverse applications, including boat flotation aids; structural parts in aircraft, submarines, and missiles; structural cores for wall panels; and ablative heat shields for reentry vehicles and rocket test engines.

Microballoons have been used for gap filling, where the spheres dampen sound or vibration in the structure. In the medical area, microballoons have been evaluated as a skin replacement for burn victims and phantom tissue for radiation studies. An important application is in nitroglycerin-based explosives, in which microballoons permit a controlled sequential detonation not possible with glass spheres.

In the 1990s, carbon microbeads have been produced by a proprietary process using phenolic resin. Potential applications are lubricants, adhesives, and conductive fillers for plastics, rubbers, and coatings (97).

Fibers. The principal type of phenolic fiber is the novoloid fiber (98). The term novoloid designates a content of at least 85 wt % of a cross-linked novolak. Novoloid fibers are sold under the trademark Kynol, and Nippon Kynol and American Kynol are exclusive licensees. Novoloid fibers are made by acid-catalyzed cross-linking of melt-spun novolak resin to form a fully cross-linked amorphous network. The fibers are infusible and insoluble, and possess physical and chemical properties that distinguish them from other fibers. Applications include a variety of flame- and chemical-resistant textiles and papers as well as composites, gaskets, and friction materials. In addition, they are precursors for carbon fibers.

The fibers are prepared from a high molecular weight novolak resin. Uncured fibers are prepared by melt-spinning the novolak. These fibers are then immersed in aqueous formaldehyde solution containing an acidic catalyst. As heat is applied, curing commences and the novolak resin is transformed into a cross-linked network through the formation of methylene and dibenzyl ether linkages. The final cross-linked structure is free of molecular orientation, and the density of cross-linking is low. The fiber contains ca 5 wt % unreacted methylol groups, which can be utilized in the formation of novoloid–fiber composites, or be reduced by heating the cured fiber to 180°C.

Optimum mechanical properties of the fibers are developed provided the precursor novolak filaments are less than 25 μm in diameter to ensure sufficient diffusion of the formaldehyde and catalyst into the fiber. The individual fibers are generally elliptical in cross section. Diameters range from 14 to 33 μm (0.2–1.0 tex or 2–10 den) and fiber lengths are 1–100 mm. Tensile strength is 0.11–0.15 N tex (1.3–1.8 g den) and elongation is in the 30–60° range. Elastic recovery is as high as 96°.

The limiting oxygen index (LOI) of novoloid materials varies with the particular structure (fiber, felt, fabric) being evaluated; it is generally in the 30–34° range. By comparison, aramid fiber has a LOI of 28–31 and wool of 24. When exposed to flame, novoloid materials do not melt but gradually char until completely carbonized. The high strength of the phenolic char results in the fiber retaining its original fiber structure, and the char effectively absorbs heat from the materials. When novoloid products are exposed to flame, the products of combustion are principally H₂O, CO₂, and carbon char. Smoke emission is minimal, less than that of any other organic fiber.

Applications for novoloid fibers include a variety of flame-resistant protective clothing, safety accessories, and flame barriers for upholstered furniture. As an asbestos replacement, novoloid fibers have been used in gasketing, packings, brake linings, and clutch facings. In electrical applications, novoloid fibers and papers can be used as flameproof coatings and as wrapping tapes in wire and cable applications.

The novoloid molecular structure includes methylol groups, which are available for cross-linking with reactive sites in the matrix material. The ability of the fibers to react with matrix resins yields synergistic improvements in the properties of the composite. Novoloid fibers have been incorporated into composites with thermoplastic resins such as polypropylene, PVC, and polyamide polyesters. Thermosetting resins include phenolics, epoxies, and melamines. In certain elastomers, the methylol reactivity can be used to upgrade the high temperature performance, for example, in chlorinated polyethylene. A process has been announced that produces carbonized Kynol fiber having extremely high surface area (99).

BIBLIOGRAPHY

"Phenolic Resins and Plastics" in *ECT* 1st ed., Vol. 10, pp. 335–370, by K. C. Frisch, General Electric Co.; "Phenolic Resins" in *ECT* 2nd ed., Vol. 15, pp. 176–208, by W. A. Keutgen, Union Carbide Corp.; in *ECT* 3rd ed., Vol. 17, pp. 384–416, by G. L. Brode, Union Carbide Corp.

1. "21st Century Phenolics: Properties, Processes, Applications," *SPI Phenolic Molding Division Technical Conference*, Cincinnati, Ohio, June 1987.
2. S. R. Sandler and W. Caro, *Phenol Aldehyde Condensation*, Vol. 1, Academic Press, Inc., New York, 1977, Chapt. 2.
3. R. W. Martin, *The Chemistry of Phenolic Resins*, John Wiley & Sons, Inc., New York, 1956.
4. N. J. L. Megson, *Phenolic Resin Chemistry*, Butterworth & Co. Ltd., Kent, U.K., 1958.
5. J. F. Walker, *Formaldehyde*, Van Nostrand Reinhold Co., Inc., New York, 1974.
6. T. S. Carswell, *Phenoplasts*, Interscience Publishers, New York, 1947.
7. A. Knop and W. Scheilb, *Chemistry and Applications of Phenolic Resins*, Springer-Verlag, Berlin, 1979.
8. A. Knop and L. Pilato, *Phenolic Resins*, Springer-Verlag, Berlin, 1985.
9. ACS Symposium on Phenolic Resins, Washington, D.C., Aug. 1983.
10. **U.S. Pat. 643,012** (Feb. 6, 1900), A. Smith.
11. **U.S. Pats. 939,966** (Nov. 15, 1909) and **942,852** (Dec. 12, 1909), L. H. Baekeland.
12. J. B. Flemming, *Hydrocarbon Process.* **55**(1), 185 (Jan. 1976).
13. K. Kamide and Y. Miyakawa, *Makromol. Chem.* **179**, 359 (1978).
14. *Hydrocarbon Process.* 90 (Mar. 1976); 216 (Nov. 1965); 183 (Nov. 1969); 150 (Nov. 1975); 135 (Nov. 1973).
15. P. Kopf and E. Wagner, *J. Polym. Sci. Polym. Chem. Ed.* **11**(5), 939 (1973).
16. Y. Ihashi, K. Sawa, and S. Morita, *Kogyo Kagaku Zasshi*, **68**, 1427 (1965).
17. **U.S. Pats. 4,433,119** and **4,433,129** (Feb. 21, 1984), S. W. Chow and G. L. Brode (to Union Carbide Corp.).
18. S. Sojka, *Macromolecules*, **12**, 767 (1979).
19. S. Ishida, Y. Tsutsumi, and K. Katsumasa, *J. Polym. Sci. Polym. Chem. Ed.* **19**, 1609 (1981).
20. H. Bender, *Ind. Eng. Chem.* **44**, 1619 (1952).
21. A. Zsavitsas and A. Beaulieu, *Am. Chem. Soc. Div. Org. Coat. Plast. Chem.* **27**, 100 (1967).
22. H. Peer, *Rec. Trav. Chim.* **78**, 851 (1959); **79**, 825 (1960).
23. J. H. Freemann and C. Lewis, *J. Am. Chem. Soc.* **76**, 2080 (1954).
24. K. Eapen and L. Yeddanapalli, *Makromol. Chem.* **4**, 119 (1968).
25. **U.S. Pat. 4,124,554** (Nov. 7, 1978), J. Fry (to Union Carbide Corp.).
26. **U.S. Pat. 4,206,095** (June 3, 1980), J. Wynsta and S. Schultz (to Union Carbide Corp.).
27. **U.S. Pat. 3,943,080** (Mar. 9, 1976), M. Wismer (to PPG Industries).
28. H. A. Bruson and C. W. MacMullen, *J. Am. Chem. Soc.* **63**, 270 (1941).
29. T. T. Jones, *J. Soc. Chem. Ind. London*, **65**, 264 (1946).
30. O. Vlk, *Plaste Kautsch.* **4**, 127 (1957).
31. **U.S. Pat. 3,687,896** (Aug. 29, 1972), S. Vargia and co-workers (to Societa Italiana Resine SpA).
32. *Mod. Plast. Int. Ed.* **58**(4), 45 (Apr. 1981).
33. *Chem. Eng.* **99**(6), 19 (June 1992); **U.S. Pat. 4,900,671** (1990), A. Pokora (to Mead Corp.); **U.S. Pat. 5,153,298** (1992), A Pokora (to Mead Corp.).
34. **U.S. Pat. 4,113,700** (1978), W. Aubertson (to Monsanto Co.).
35. **U.S. Pat. 3,823,103** (July 9, 1979), J. Harding (to Union Carbide Corp.).
36. G. Brode, T. Jones, and S. Chow, *Chem. Technol.* **13**, 676 (1983).
37. **U.S. Pat. 4,366,303** (Dec. 28, 1982), P. W. Kopf (to Union Carbide Corp.).
38. **U.S. Pat. 4,788,236** (Nov. 29, 1988), P. W. Kopf (to Union Carbide Corp.).
39. **U.S. Pat. 2,797,201** (July 11, 1957), F. Veatch (to Sohio).
40. **U.S. Pat. 4,098,770** (July 4, 1978), A. Berchem, K. Sudan, and E. Gres (to Reichhold Chemicals, Ltd., Canada).
41. K. Hultzsche, *Kunststoffe*, **37**, 205 (1947).
42. E. Orrell and R. Burns, *Plast. Polym.* **68**, 469 (Oct. 1968).
43. V. Pshenitsyna and M. Kotreleev, *Polym. Sci. USSR* **11**, 829 (1969).
44. V. Kurachenkov and L. Igonin, *J. Polym. Sci. Part A* **1**(9), 2283 (1971).
45. M. Yamao, N. Yoshimi, and S. Tanaka, *Kogyo Kagaku Zasshi* **74**, 1938 (1971).
46. J. Gillham, *Polym. Prepr. Am. Chem. Soc. Div. Polym. Chem.* **22**(2), 131 (1981).
47. R. Young, P. Kopf, and O. Salgado, *Tappi*, **64**(4), 127 (1981).
48. H. L. Bender, *Mod. Plast.* **30**, 136 (1953).
49. Y. Zahs and co-workers, *J. Appl. Polym. Sci.* **27**, 913 (1982).
50. J. Woodbrey, H. Higginbottom, and H. Culbertson, *J. Polym. Sci. Part A* **3** 1079 (1965).
51. R. Hirst and co-workers, *J. Polym. Sci. Part A* **3**, 2091 (1965).
52. T. Yoshikawa and J. Kumanotani, *Makromol. Chem.* **131**, 273 (1970).
53. B. Mechin and co-workers, *Eur. Polym. J.* **20**, 333 (1984).
54. J. Shaefer, *J. Am. Chem. Soc.* **98**, 1031 (1976).
55. R. Bryson and co-workers, *Macromolecules* **16**, 1669 (1983).

56. C. Fyfe, *Macromolecules* **16**, 1216 (1983).

57. G. Maciel and co-workers, *Macromolecules* **17**, 1081 (1984).

58. E. Wagner and R. Greff, *J. Polym. Sci. Part A* **1**(9), 2193 (1971).

59. D. Braun and J. Arnelt, *Angew. Makromol. Chem.* **73**, 133 (1978).

60. A. Rudin, C. Fyfe, and S. Vines, *J. Appl. Polym. Sci.* **28**, 2611 (1983).

61. E. Tesarova and V. Pacakova, *Chromatographia* **17**(5), 269 (1983).

62. W. Werner and O. Barber, *Chromatographia*, **15**, 101 (1982).

63. L. Taylor, *Anal. Chem.* **42**, 297 (1970).

64. M. Kamal and S. Sourour, *Polym. Eng.* **13**, 59 (1973).

65. J. Enns and J. Gillham, *J. Appl. Polym. Sci.* **28**, 2567 (1983).

66. H. Babich and D. Davies, *Reg. Tox. Pharm.* **1**(1), 90 (1981).

67. Code of Federal Regulations, 1910.1048 Title 29, 1989.

68. T. Sterling and J. Weinkam, *J. Occupation. Med.* **30**(11), 895–901 (Nov. 1988).

69. A. Housden, *Water Pollut. Control* **80**(4), 490 (1981).

70. *Reinforced Plast.* **35**(7,8) 17 (1992).

71. *Plast. Week* **2** (Mar. 23, 1992).

72. *Plast. Environ.* **5** (July 10, 1992).

73. "SPI Committee of Resin Statistics," *Mod. Plast.* **71**(1), 73 (1994).

74. *Society of Plastic Industries Facts and Figures*, SPI, Washington, D.C., 1994.

75. G. Bosshard, *Prog. Org. Coat.* **10**, 205 (1982).

76. **U.S. Pats. 3,723,156** (Mar. 27, 1973) and **3,732,120** (May 8, 1973), B. Brockett; **3,737,410** (June 5, 1973), H. Mueller; **3,936,573** (Feb. 3, 1976), B. Brockett; and **4,165,102** (Aug. 21, 1979), J. Bodmer (to National Cash Register).

77. W. A. Boyd, C. Merriam, and J. Fry, in R. Tess and G. Behlein, eds., *ACS Symposium Series, Applied Polymer Science*, 2nd ed., 1985, pp. 1141–1158.

78. D. M. Berger, *Met. Finish.* **74**(9), 52 (1976).

79. J. Fry, C. Merriam, and G. Misko, *Am. Chem. Soc. Div. Org. Coat. Plast. Chem.* **47**, 540 (1982).

80. B. Barth, in E. Skeist, ed., *Handbook of Adhesive*, Van Nostrand Reinhold Co., Inc., New York, 1977, Chapt. 23, p. 382.

81. **U.S. Pat. 3,595,821** (July 27, 1971), S. Spector (to PPG Industries).

82. H. Kollek, H. Brockmann, and H. Mueller, *Int. J. Adhesion Adhes.* **6**(1), 37 (1986).

83. **U.S. Pat. 4,025,490** (May 24, 1977), F. Weaver (to Mead Corp.).

84. J. Perry, *Pulp Pap.* **46**, 93 (Aug. 1973); 116 (Oct. 1973).

85. G. Elbe, *Plastics Design Forum*, 66 (Nov. Dec. 1993).

86. **U.S. Pats. 4,316,827** (Feb. 23, 1982) and 4,420,571 (Dec. 13, 1983), L. Pacala and J. Blickensderfer (to Union Carbide Corp.).

87. **U.S. Pat. 4,403,066** (Sept. 6, 1983), G. L. Brode, S. Chow, and W. Hale (to Union Carbide Corp.).

88. *ACS National Meeting*, Washington, D.C., Aug. 1983; G. L. Brode, *Polym. Prepr.* **24**(2), 194 (1983).

89. **U.S. Pat. 4,419,400** (Dec. 6, 1983), R. Hindersinn (to Occidental Petroleum).

90. *Plast. World*, **47**(6), 42 (1989).

91. J. Delmonte, *Technology of Carbon and Graphite Fiber Composites*, Van Nostrand Reinhold Co., Inc., New York, 1981.

92. E. F'itzer and W. Schafer, *Carbon*, **8**, 353 (1970).

93. S. Siebold, *SAMPE Symp.* **20**, 327 (Apr. 1975).

94. **U.S. Pat. 4,395,521** (July 26, 1983), S. W. Chow (to Union Carbide Corp.).

95. G. L. Brode, S. W. Chow, and M. Michno, *Polym. Prepr. Am. Chem. Soc. Div. Polym. Chem.* **24**(2), 192 (Aug. 1983).

96. **U.S. Pats. 4,423,163** (Dec. 27, 1983), H. Doerge; **4,165,413** (Aug. 21, 1979), R. Sefton; and **444,912** (Apr. 24, 1984), J. Carlson (to Koppers Co., Inc.).

97. *J. Chem. Week*, **30**(1526) 4 (1989).

98. **U.S. Pats. 3,650,102** (Mar. 21, 1972) and **3,723,588** (Mar. 27, 1973), J. Economy and R. Clark (to Carborundum Co.).

99. *J. Chem. Week*, **5** (Aug. 10, 1989).

General References

E. Dradi and G. Casiraghi, *Macromolecules* **11**, 1295 (1978).

T. Liu and S. Rhee, *W'ear* **76**, 213 (1978).

Peter W. Kopf
Arthur D. Little, Inc.

PHENOLSULFONIC ACIDS.

See SULFONIC ACIDS.

PHENOTHIAZINE.

See AZINE DYES; PSYCHOPHARMACOLOGICAL AGENTS.

PHENYLACETIC ACID.

See BENZYL ALCOHOL AND β-PHENETHYL ALCOHOL; PERFUMES.

PHENYLENEDIAMINES AND TOLUENEDIAMINES.

See AMINES, AROMATIC—PHENYLENEDIAMINES; AMINES, AROMATIC—DIAMINOTOLUENES.

PHEROMONES.

See INSECT CONTROL TECHNOLOGY; HORMONES, SURVEY.

PHORATE.

See INSECT CONTROL TECHNOLOGY.

PHOSGENE

PhosgenePhosgene [75-44-5] [75-44-5] (carbonyl chloride, carbon oxychloride, chloroformyl chloride), Cl₂CO, is a colorless, low boiling liquid. The compound was first prepared in 1812 by J. Davy from the photochemical reaction of carbon monoxide and chlorine. Phosgene may be formed at elevated temperatures by oxidation of chlorinated solvents (1–5). Aside from its use as a warfare agent in World War I, phosgene has been used in the preparation of a great variety of chemical intermediates. It is widely used in the preparation of isocyanates which are used in the preparation of polyurethanes (see URETHANE POLYMERS), in the manufacture of polycarbonate, and in the synthesis of chloroformates and carbonates which are used as intermediates in the synthesis of pharmaceuticals and pesticides (see CARBONIC AND CARBONCHLORIDIC ESTERS). Because of its toxicity, a high level of safety technology has been developed to help ensure the safe handling of phosgene.

Properties

Some physical properties of phosgene are listed in Table 1. At room temperature and normal pressure, it is a colorless gas. Impurities may cause discoloration of the product to pale yellow to green. Phosgene has a characteristic odor; the odor of the gas can be detected only briefly at the time of initial exposure. At ca 0.5 ppm in air, the odor has been described as pleasant and similar to that of new-mown hay or cut green corn. At high concentration, the odor may be strong, stifling, and unpleasant. In general, phosgene is soluble in aromatic and aliphatic hydrocarbons, chlorinated hydrocarbons, and organic acids and esters. It is removed easily from solvents by heating or sparging with air or nitrogen, but because of its toxicity, great care must be taken to prevent its presence in the atmosphere.

Table 1. Some Physical Properties of Phosgene ^a

Properties and characteristics	Value
molecular weight	98.92
melting point, °C	127.84
boiling point, °C ^b	7.48
density at 20°C, g cm ³	1.387
vapor pressure at 20°C, kPa ^c	161.68
vapor density (air = 1.0)	3.4
critical temperature, °C	182
density at critical point, g cm ³	0.52
critical pressure, MPa ^d	5.68
latent heat of vaporization, at 7.5°C, J g ^e	243

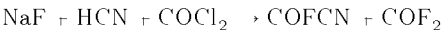
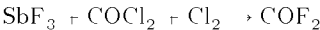
molar heat capacity of liquid, at 7.5°C, J K [°]	100.8
molar heat of formation, kJ [°]	
from elements	218
from CO and Cl ₂	108
molar entropy, J K [°]	
at 7°C	280
25°C	284
surface tension, mN m(dyn cm)	
at 0.0°C	34.6
16.7°C	20.1
34.5°C	17.6
46.1°C	15.9

^a Ref. 6.
^b At 101.3 kPa^c = 1 atm.
^c To convert kPa to psi, multiply by 0.145.
^d To convert MPa to psi, multiply by 145.
^e To convert J to cal, divide by 4.184.

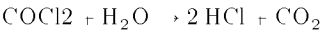
Phosgene is a planar molecule. The interatomic distances are 0.128 nm for the C–O bond, and 0.168 nm for C–Cl (1). The angle Cl–C–Cl is 117°. Infrared, ultraviolet, and Raman spectral properties have been described (7–9).



Reactions. Phosgene interacts with many classes of inorganic and organic reagents. The reactions have been described extensively (10). Reaction with sodium metal takes place at room temperature, but reaction with zinc requires warming. Phosgene is an excellent chlorinating agent. Oxides and sulfides of metals react with phosgene at elevated temperatures, yielding usually very pure chlorides. The reaction of phosgene with cadmium sulfide is a good method for preparing carbonyl sulfide (carbon oxysulfide), COS. The reactions of phosgene with oxides of calcium, magnesium, tin, titanium, tungsten, and zinc have been described (11–18). Chlorination with phosgene of oxides of interest in the nuclear field, especially uranium oxide, plutonium oxide, and thorium oxide, also has been investigated (19–22). Phosphates and silicates of metals often react with phosgene at elevated temperatures and yield the metal chloride and phosphorus oxychloride or silicon dioxide. The reaction with ferric phosphate at 300–350°C has been proposed as a synthetic method for phosphorus oxychloride, POCl₃. Anhydrous aluminum chloride forms a variety of complexes with phosgene, eg, Al₂Cl₃·5COCl₂ at low temperatures, Al₂Cl₃·3COCl₂ at 30°C, and Al₂Cl₃·COCl₂ above 55°C. Reaction with aluminum bromide yields carbonyl bromide [593-95-3], COBr₂, and aluminum chlorobromide, AlCl₂Br. Antimony trifluoride with phosgene and chlorine yields carbonyl fluoride (23). Phosgene reacts with sodium fluoride and HCN, yielding phosgene fluorocyanide and carbonyl fluoride [353-50-4] (24).

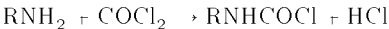


Phosgene reacts slowly with cold water to give CO₂ and HCl and more quickly at higher temperatures (25). In the reaction of gaseous phosgene with water, it is difficult to get the necessary intimate mixing of the gas and water.

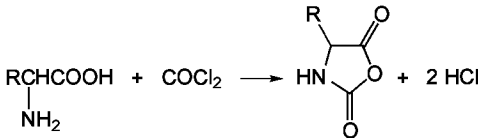


Ammonia reacts vigorously with phosgene. The products are urea, biuret, ammelide (a polymer of urea), cyanuric acid, and sometimes cyamelide (a polymer of cyanic acid). The secondary products probably arise through the very reactive intermediate carbamyl chloride [463-72-9], NH₂COCl (see CYANAMIDES).

Phosgene reacts with a multitude of nitrogen, oxygen, sulfur, and carbon centers. Reaction with primary alkyl and aryl amines yield carbamoyl chlorides which are readily dehydrohalogenated to isocyanates. Secondary amines also form carbamoyl chlorides.



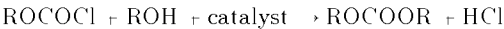
α-Amino acids react readily with phosgene to form oxazolidine-2,5-diones:



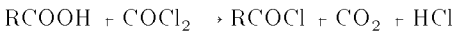
Hydrazine reacts with phosgene yielding carbohydrazide:



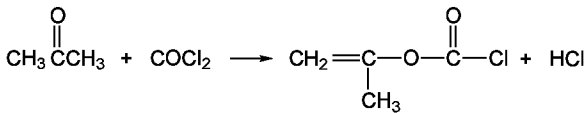
The reaction of phosgene with alcohols yields chloroformates, and with a basic catalyst present, carbonates are formed:



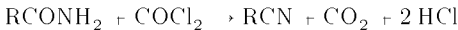
This reaction is commercially important because it serves as a basis for the manufacture of polycarbonate. Carboxylic acids react with phosgene to give acid chlorides (26) (see CARBOXYLIC ACIDS).



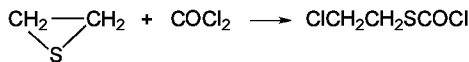
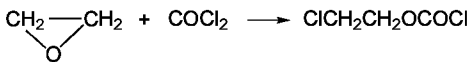
Ketones also react with phosgene:



Amides react with phosgene to yield nitriles (qv).



Phosgene also can initiate ring opening:



Although POCl₃ is the traditional reagent in the Vilsmeier aldehyde synthesis, phosgene may be employed (27–29).

Manufacture

Phosgene is manufactured by reaction of carbon monoxide with chlorine over activated carbon. Depending on the quantity needed and availability of the raw materials, numerous variations of the basic synthetic process are being practiced. Continuous processing and a high degree of automation are required for phosgene purification, condensation, and storage. Because of its toxicity, careful and extensive safety procedures and safety equipment are incorporated in plant design and operation. The manufacture of phosgene consists of preparation and purification of carbon monoxide, preparation and purification of chlorine, metering and mixing of reactants, reaction of mixed gases over activated carbon catalyst, purification and condensation of phosgene, and recovery of traces of phosgene to assure worker and environmental safety.

Carbon monoxide (qv) may be manufactured according to standard processes from coke, ie, from coal, or by controlled oxidation of hydrocarbon fuels. A carbon monoxide process must be chosen that yields a gas of the highest possible purity. Noncondensable impurities are particularly objectionable since their presence makes the recovery of phosgene difficult. Water must be removed from the starting gas to preclude hydrochloric acid formation in the phosgene generator. The hydrocarbon and hydrogen content should be minimized because reaction of chlorine with methane or hydrogen could ignite a reaction between chlorine and steel, thereby destroying the equipment. Other impurities might poison the activated carbon catalyst. Sulfides must be excluded since they produce sulfur chlorides which usually are undesirable impurities. The chlorine must be as dry and pure as the carbon monoxide to avoid corrosion of the equipment and decomposition of phosgene by water and other impurities (see CORROSION AND CORROSION CONTROL; HYDROCARBONS).

Activated carbon of high absorptive capacity is suitable for use as a catalyst; it need not be treated with metallic salt or other substances. If starting materials of high purity are employed, excellent and economic catalyst efficiency is obtained.

The phosgene generators employed are relatively simple, tubular heat exchangers that are filled with granulated activated carbon (see CARBON, ACTIVATED CARBON). Because the reaction is rapid and exothermic, efficient heat removal is important. Decomposition of phosgene into its starting materials begins to take place at 200°C. The temperature of the carbon bed in the initial reaction zone of the tubes can reach 400°C, but it rapidly falls to product temperatures of 40–150°C. The reaction is generally run at normal pressure or at a slight excess pressure. A phosgene generation system which monitors the phosgene requirements of a plant and responds by producing only the needed amounts has been developed (30,31).

A flow diagram of the production of phosgene is given in Figure 1. Carbon monoxide and chlorine gas are mixed in equimolecular proportions. A small amount of excess carbon monoxide may be used to ensure complete reaction of the chlorine. The product gases can be condensed leading to liquid phosgene and uncondensed gases, which are then scrubbed for removal of remaining phosgene. Uncondensed gaseous phosgene can be employed for in-line operations. In some cases, phosgene is absorbed into a solvent in a so-called absorption column. The solvent used for the absorption is typically the solvent used in a later process step. The remaining nonabsorbable gas stream is fed to the waste gas treatment system to be freed from phosgene.

Waste Gas Streams. Several methods of decomposing phosgene in waste gas streams are used. The outlet gas from the phosgene decomposition equipment is continuously monitored for residual phosgene content to ensure complete decomposition.

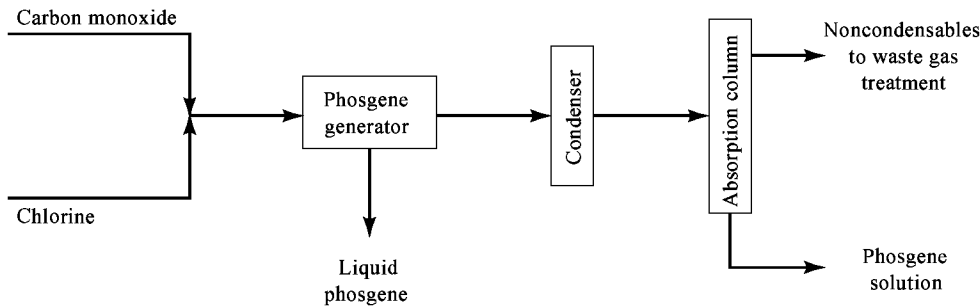


Fig. 1. Manufacture of phosgene from carbon monoxide and chlorine.

Decomposition by Caustic Scrubbing. The waste gas stream is led through packed towers where a sodium hydroxide solution is introduced at the top of the towers. Venturi scrubbers can also be used. Makeup sodium hydroxide is added under pH control (32).

Decomposition with Moist Activated Carbon. The waste gas stream is passed through packed activated carbon towers where water is fed at the top of the towers. The water is normally recycled. If the hydrochloric acid concentration in the recycled water exceeds 10⁰ ‰, the decomposition efficiency is greatly reduced. Thus, a sufficient supply of fresh water must be assured and a hydrochloric acid stream continuously taken out (33).

Combustion. The waste gas stream is burnt to convert phosgene to carbon dioxide and HCl. An advantage of this method is that all components of the waste gas, such as CO and solvent, are burnt (34).

Analytical and Test Methods

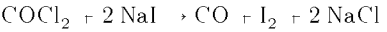
Phosgene in air and in mixture with other gases can be detected by a variety of methods (35). Trace quantities to a lower limit of 0.05 µg L air can be detected by uv spectroscopy (36). Both ir and gas chromatography have been used extensively to measure phosgene in air at 1 ppb–1 ppm (7,37,38). Special

and multiple-column gas-chromatographic methods have been used for more complex mixtures of gases containing phosgene (39–41). High performance liquid chromatography (hplc) methods can also be used and offer detection limits of 5–10 ppb phosgene (42,43). Absolute determination of phosgene at levels below 100 ppb has been reported using pulsed flow coulometry (44). Laser photoacoustic spectroscopy has been used to detect phosgene at ppb levels (45) (see LASERS).

Methods and instruments that are used to monitor phosgene content in air are well developed and have been reviewed (46–48). One detection instrument is a porous tape that measures the concentration of phosgene in air in quantities as small as 6 ppb (49). Fourier transform ir spectrometry techniques have been developed to permit line and area monitoring in the area around phosgene plants (50).

A phosgene dose-indicator badge for personnel exposure monitoring has been prepared (51). This is a simple, visually readable, passive sampler based on phosgene indicator paper which is impregnated with a solution of 4-*p*nitrobenzylpyridine and *N*-benzylaniline. These indicator badges are used at a number of plant sites using phosgene.

Liquid phosgene is assayed by an iodometric method which involves the following reaction (52). The released iodine is titrated with sodium thiosulfate.



The following specifications and standards have been reported (6):

Assay	Percent
COCl ₂ , min	99.0
Cl ₂ (free), max	0.1
HCl, max	0.2

Storage and Handling

All phosgene containers require a Class A, poison gas label as well as a corrosive label. Phosgene is transported in steel cylinders which conform to rigid safety design specifications. The cylinders undergo special hydrostatic testing at 5.5 MPa (800 psi), and extension rings are incorporated in the cylinders to protect the valves. Phosgene is shipped in cylinders ranging in size from 43 to 909 kg. Careful testing for leaks is required after filling, and a vapor space must be accommodated in the storage vessel; excessive filling with liquid phosgene must be avoided.

Transportation requirements and classifications for phosgene are as follows:

DOT shipping name	phosgene
DOT hazard class	2.3
Reportable quantity (RQ)	yes - 10 lbs (4.5 kg)
DOT labels	poison, corrosive
DOT placards	poison
bill of lading description	phosgene, 2.3, UN1076, RQ (phosgene), inhalation hazard, poison gas and corrosive labels affixed
UN NA number	UN1076
additional DOT requirements	return of empty containers: residue last contained: phosgene, 2.3, UN1076, RQ (phosgene), inhalation hazard, poison gas and corrosive labels affixed

Because phosgene reacts with water, great care must be taken to prevent contamination with traces of water since this could lead to the development of pressure by hydrogen chloride and carbon dioxide. Wet phosgene is very corrosive; therefore phosgene should never be stored with any quantity of water (4).

Health and Safety Factors

The odor threshold for phosgene is ca 0.5–1 ppm, but it varies with individuals and is higher after prolonged exposure (53). Phosgene may irritate eyes, nose, and throat. The permissible exposure TLV by volume in air is 0.1 ppm (54). The TLV refers to the average airborne concentration at which it is believed nearly all workers may be repeatedly exposed on a daily basis without adverse effect. It is a time-weighted average for an 8-h day or a 40-h week and should be used as a guide for control only. The guideline for excursion limits above the TLV is 0.2 ppm (55). Long-term exposure to phosgene has been reviewed, and potential hazards may exist at concentrations slightly higher than the TLV (56). Medical problems and adverse health effects associated with phosgene exposure have been reviewed (57–59), and therapy for phosgene poisoning has also been reviewed (60).

When phosgene is inhaled, it reacts very little with the aqueous film on the mucous membranes of the upper respiratory tract. Most of it reaches the pulmonary alveoli, where gas exchange occurs. Here phosgene reacts with NH₂, OH, and SH groups of the blood–air barrier. This causes the blood–air barrier to lose its function as a membrane between the blood vessels and the pulmonary alveoli. As a result, blood plasma is able to pass from the blood vessels into the pulmonary alveoli, thereby increasingly disturbing the gas exchange, ie, the normal lung function.

Breathing phosgene causes pulmonary edema which may be characterized by a delayed onset. Exposed persons must be removed immediately from the contaminated area. Rescue workers should wear self-contained breathing apparatus. Injured persons should not be allowed any physical activity, and a physician should be consulted immediately. When breathing is occurring or has been restored, oxygen should be administered. Oxygen should be given as long as necessary to maintain normal color of the skin. The patient should be kept comfortably warm but not hot. It is advisable to keep an exposed individual under the observation of a physician for 6–24 hours, depending on the circumstances of exposure. In some instances this can be accomplished at the plant medical facility, or the exposed individual may require hospitalization.

In handling phosgene, extensive safety precautions and procedures are required to prevent exposure to phosgene. The first point is to design the phosgene system to prevent phosgene emissions from the closed equipment. Chemically resistant, high quality materials are used for plant equipment and lines which are inspected regularly. Pumps, which are hermetically sealed to the outside, are used for phosgene-containing streams. Stringent requirements are set for the quality and design of phosgene process control equipment, and critical areas have redundant control systems. The second point is to quickly detect leaks and contain or decompose escaped phosgene. This includes such measures as continuous alarm systems to monitor the working atmosphere, systems for decomposing escaped phosgene (eg, steam–ammonia curtains for gaseous emissions), jacketed pipes, and complete containments for phosgene plants (61,62).

In case of extensive leaks or spills, immediate evacuation upwind of the phosgene source is necessary. Phosgene is 3.4 times as heavy as air and may collect in low lying areas (6). Water should not be used on the source of a phosgene leak because the resulting corrosion enlarges the leak. Suitable personal protective equipment includes respiratory equipment and eye protection. In case of fire, it is essential to cool all phosgene-containing vessels. Leaks of liquid phosgene or phosgene solutions can be effectively combated by covering the phosgene-containing liquid with an absorbent and decomposition agents.

Waste Disposal. Because of its low boiling point and high toxicity, measures must be taken to prevent the entrance of phosgene into drains or sewers. If recycle of phosgene is not feasible, phosgene waste can be handled by one of the decomposition methods mentioned above, ie, caustic scrubbing, moist activated carbon towers, or combustion.

Uses

Phosgene is an important and widely used intermediate. Practically all phosgene manufacture is captive, ie, it is used in the manufacture of other chemicals within the plant boundary. Only VanDeMark sells phosgene on the U.S. merchant market. The U.S. demand and use of phosgene for 1994 was 1.18×10^6 metric tons, and for 1995, it is projected to grow to 1.23×10^6 metric tons (63). The primary use of phosgene is the manufacture of polyisocyanates for the polyurethane industry, where over 80% of the phosgene output in the United States is consumed. The manufacture of toluene diisocyanate [1321-38-6] accounts for ca 45% of the phosgene consumption, whereas diisocyanato diphenylmethane [101-68-8] (MDI) and polymeric MDI [9016-87-9] account for ca 35% (see AMINES, AROMATIC—METHYLENEDIANILINE). The polycarbonate industry accounts for ca 10% of the phosgene consumption, and the remaining 10% is used to produce aliphatic diisocyanates, monoisocyanates, chloroformates, agrochemicals, and intermediates for pharmaceuticals and dyestuffs. Although several commercial phosgene-free processes for producing polyisocyanates and polycarbonates have been announced, they are not expected to make large inroads into phosgene demand in the 1990s (64–68).

Phosgene can be employed in a variety of metal-recovery operations, eg, in the recovery of platinum, uranium, plutonium, and niobium (69–73). Phosgene has been proposed for the manufacture of aluminum chloride, beryllium chloride, and boron trichloride (74–76). Phosgene has been patented as a stabilizer, either by itself or in combination with thionyl chloride, for liquid SO₂ (77).

Phosgene also is used in the production of chloroformates which are intermediates in the production of ore-flotation agents, perfumes (qv), herbicides (qv), and insecticides (see INSECT CONTROL TECHNOLOGY), and especially in the preparation of pharmaceuticals (qv) (78). Chloroalkyl chlorocarbonates have been manufactured as intermediates for pharmaceuticals and pesticides (qv) (79,80). Phosgene also is employed in the manufacture of carbonic esters. Phosgene has been suggested as a starting material for the manufacture of carbon tetrachloride and other chlorinated hydrocarbons (81–85). A number of pesticides have been patented based on the reaction of a thiol or dithiol with phosgene to form thiol chloroformates (86,87). The preparation of aromatic acids using phosgene, and a detailed process for the manufacture of terephthalic acid from toluene and phosgene, have been described (88,89) (see PHTHALIC ACID AND OTHER BENZENEPOLYCARBOXYLIC ACIDS).

An important direct use of phosgene is in the preparation of polymers. Polycarbonate is the most significant and commercially valuable material (see POLYCARBONATES). However, the use of phosgene has been described for other polymer systems, eg, fiber-forming polymeric polyketones and polyureas (90,91).

BIBLIOGRAPHY

"Phosgene" in *ECT* 3rd ed., Vol. 13, pp. 416–425, by E. E. Hardy, California State Polytechnic University.

1. L. C. Riusema and L. G. Silverstein, *Am. Ind. Hyg. Assoc. J.* **33**(1), 35 (1972).
2. M. H. Noweir, E. A. Pfitzer, and T. F. Hatch, *Am. Ind. Hyg. Assoc. J.* **34**(1), 25 (1973).
3. H. F. Anderson, J. A. Dahlberg, and R. Wettstrom, *Ann. Occup. Hyg.* **18**, 129 (1975).
4. H. B. Singh, *Nature* **264**, 428 (1976).
5. M. J. Thomson, B. S. Biggins, D. Lucas, C. P. Koshland, and R. F. Sawyer, *Combust. Flame* **98**, 350 (1994).
6. *Chemical Safety, Data Sheet SD-95*, Manufacturing Chemists Association, Washington, D.C., rev. 1978.
7. G. G. Esposito, *Anal. Chem.* **49**, 1774 (Oct. 1977).
8. V. Henri and O. R. Howell, *Proc. Roy. Soc. (London) Ser. A* **128**, 190 (1930).
9. R. Anathakrishnan, *Proc. Indian Acad. Sci. Sect. A* **5**, 285 (1937).
10. H. Babad and A. G. Zieler, *Chem. Rev.* **73**(1), 75 (1973).
11. A. N. Ketov, V. V. Pechkovskii, and L. P. Kostin, *Issled. Obl. Khim. Tekhnol. Mineral's Solei Okislov, Akad. Nauk SSR Sb. Statei*, 202 (1965).
12. A. N. Ketov, V. V. Pechkovskii, and L. P. Kostin, *Izv. Vyssh. Uchebn. Zaved. Tsvetn. Met.* **7**(2), 94 (1964).
13. A. N. Ketov, V. V. Pechkovskii, and L. P. Kostin, *Zh. Neorg. Khim.* **9**, 467 (1964).
14. Yu. P. Kuznetsov, E. S. Petrov, and A. I. Vakhrusheva, *Izv. Sib. Otd. Akad. Nauk SSR Ser. Khim. Nauk* **1**(2), 63 (1969).
15. *Ibid.*, *Izv. Sibirsk. Otd. Akad. Nauk SSR Ser. Khim. Nauk* (3), 60 (1969).
16. A. N. Ketov, V. V. Pechkovskii, and L. P. Kostin, *Sb. Nauchn. Tr. Permsk. Politekhn. Inst.* (14), 3 (1963).
17. A. N. Ketov, V. V. Pechkovskii, and I. M. Kolesov, *Izv. Vyssh. Uchebn. Zaved. Khim. Khim. Tekhnol.* **9**, 570 (1967).
18. A. N. Ketov and I. M. Kolesov, *Sb. Nauchn. Tr. Permsk. Politekhn. Inst.* **18**, 42 (1965).
19. D. Naumann, *Kernegie* **5**(2), 188 (1962).
20. R. J. Sorenson, *U.S. at. Energy Comm.* (HW-79141), 22 (1963).
21. D. T. Peterson and D. J. Sundquist, *U.S. At. Energy Comm.* (IS-917), 30 (1964).
22. I. S. Pap, G. Mink, T. Szekely, and A. Auroux, *J. Therm. Anal.* **36**(5), 1765 (1990).
23. R. Hazeldine and H. Iserson, *J. Am. Chem. Soc.* **79**, 5801 (1957).
24. C. W. Tullock and D. D. Coffman, *J. Org. Chem.* **25**, 2016 (1960).
25. R. Mertens, C. von Sonntag, J. Lind, and G. Merenyi, *Angew. Chem.* **106**(12), 1320 (1994).
26. **Ger. Pat. 4,028,774** (Dec. 3, 1992), P. Ksoll, W. Ruether, A. Hohmann, and P. Wittmer (to BASF AG).
27. G. Martin and M. Martin, *Bull. Soc. Chim. Fr.* (8–9), 637 (1963).
28. H. H. Bosshard and H. Zollinger, *Helv. Chim. Acta* **42**, 1659 (1959).
29. *Chem. Ind.*, 682 (Sept. 21, 1992).
30. U. Osterwalder, *Ber. Int. Kolloq. Verhuetung Arbeitsunfaellen Berufskrankrankh. Chem. Ind.* **10**, 785 (1985).
31. *Chem. Eng.*, 27 (Nov. 10, 1986).
32. S. Kistner and co-workers, *Proc. Annual Meet.—Air Pollut. Control Assoc. 70th*, **2**(36) (1977).
33. **Ger. Pat. 961,681** (1953), H. Wollthan and W. Groves (to Bayer AG).
34. K. Hess and R. Kotkamp, *Chem. Ing. Tech.* **45**, 873 (1973).
35. P. C. Nigam and S. Prasad, *Indian J. Environ. Health* **28**(3), 218 (1986).
36. W. B. Crummett and J. D. Melean, *Anal. Chem.* **37**, 424 (1965).
37. B. B. Baker, *Am. Ind. Hyg. Assoc. J.* **35**, 735 (1974).
38. D. Reichert, U. Splengler, and D. J. Henschler, *Chromatog.* **179**(1), 181 (1979).
39. R. J. Graham and F. D. Stevenson, *J. Chromatog.* **47**, 555 (1970).
40. P. K. Basu, C. J. King, and S. Lynn, *J. Chromat. Sci.* **10**, 419 (1972).
41. H. Geisser and W. Richarz, *J. Chromatog.* **147**, 453 (1978).
42. W. S. Wu and S. Virindar, *Analyst* **118**(10), 1285 (1993).
43. R. J. Rando, H. G. Poovey, and S. N. Chang, *J. Liq. Chromatogr.* **16**(15), 3291 (1993).
44. H. B. Singh, D. Lillian, and A. Appleby, *Anal. Chem.* **47**, 860 (1975).
45. X. Luo, F. Y. Shi, and J. X. Lin, *J. Infrared Millimeter Waves* **12**(2), 141 (1991).

46. M. H. Noweir and E. A. Pfitzer, *Am. Ind. Hyg. Assoc. J.* **32**(3), 163 (1971).

47. B. W. Thomas, *Oil Gas J.* **63**(24), 119, 121 (1965).

48. R. M. Tuggle and co-workers, *Ind. Hyg. Assoc. J.* **40**, 387 (1979).

49. N. Nakano, A. Yamamoto, and Y. Kobayashi, *Talanta* **42**(4), 641 (1995).

50. R. H. Hunt, *Optical Remote Sensing and Applications to Environmental and Industrial Safety Problems, Apr. 6–8, 1992*, Air and Waste Management Association, Pittsburgh, Pa., 1992, p. 446.

51. W. Diller, E. Drope, and E. Reichold, *Ber. Int. Kolloq. Verbuetung Arbeit. Ber. Chem. Ind.* **6**, 137 (1979).

52. C. A. Rush and C. E. Danner, *Anal. Chem.* **20**, 644 (1948).

53. *J. Air Pollut. Control Assoc.* **19**, 91 (1969).

54. *Threshold Limit Values for Chemical Substances and Physical Agents*, American Conference of Governmental Industrial Hygienists, Washington, D.C., 1994.

55. *Phosgene Criteria Document*, NIOSH 76-137, Washington, D.C.

56. S. A. Cucinell, *Arch. Environ. Health* **28**, 272 (May 1974).

57. W. F. Diller, *J. Occup. Health* **20**, 189 (Mar. 1978).

58. W. F. Diller, *Toxicol. Ind. Health* **1**(2), 7 (1985).

59. Y. L. Yue and co-workers, *J. Appl. Physiol.* **69**(5), 1615 (1990).

60. W. F. Diller and R. Zante, *Toxicol. Ind. Health* **1**(2), 117 (1985).

61. *Chem. Week*, 34 (May 28, 1986).

62. *Chem. Eng.* **99**(5), 5 (May 1992).

63. *Chem. Mark. Rep.*, SR22 (Feb. 7, 1994).

64. S. Fukuoka, M. Chono, and M. Kohno, *J. Org. Chem.* **49**, 1458 (1984).

65. **U.S. Pat. 4,146,727** (1979), E. T. Shaw and J. G. Zajane (to Arco).

66. K. A. Henderson, Jr. and V. A. Alexanian, *Org. Prep. Proced. Int.* **18**, 149 (1986).

67. **U.S. Pat. 5,189,205** (Feb. 23, 1993), W. D. McGhee and T. E. Waldman (to Monsanto Co.).

68. W. McGhee, D. Riley, K. Christ, Y. Pan, and B. Parnas, *J. Org. Chem.* **60**, 2820 (1995).

69. **Ger. Pat. 1,077,642** (Mar. 17, 1960), G. R. Bond, Jr. (to Houdry Process Corp.).

70. **U.S. Pat. 2,890,099** (June 6, 1959), H. B. Rhodes, W. F. Pesold, and J. M. Hirshon (to U.S. Atomic Energy Commission).

71. Fr. Pat. 1,472,438 (Mar. 18, 1965), (to Atomic Energy Research Institute).

72. H. L. Brandt, AEC *Accession No. 42455*, Rpt. No. HW-83235, CFSTI, Washington, D.C., 1964, p. 20.

73. **U.S. Pat. 3,212,847** (Oct. 19, 1965), B. J. Lerner (to Dominion Gulf Co.).

74. J. Hille and W. Dьrrwachter, *Angew. Chem.* **72**, 850 (1960).

75. **U.S. Pat. 3,146,065** (Aug. 25, 1964), R. O. Bach (to Beryllium Metals and Chemical Corp.).

76. R. K. Pearson and co-workers, CCC-1024-TR-234, U.S. Atomic Energy Commission, Washington, D.C., 1957.

77. **U.S. Pat. 3,042,490** (July 3, 1960), T. W. Saults and J. J. Wimbedy (to Tennessee Corp.); **Brit. Pat. 975,318** (Nov. 18, 1964). R. V. Riley, T. F. Eden, and W. D. Hopkinson (to Stavely Iron & Chemicals Co., Ltd.).

78. **U.S. Pat. 3,152,167** (Oct. 6, 1964), M. Sletzinger (to Merck & Co., Inc.).

79. J. V. Swintosky and co-workers, *J. Pharm. Sci.* **55**, 992 (1966).

80. **Ger. Pat. 1,179,922** (Oct. 22, 1964), H. Grassner and F. Stolp (to Badische Anilin & Soda Fabrik AG).

81. R. N. Haszeldine and H. Iserson, *J. Am. Chem. Soc.* **79**, 5801 (1957).

82. O. Glemser and co-workers, *Angew. Chem.* **75**, 823 (1963).

83. **Ger. Pat. 1,188,570** (Mar. 11, 1965), W. Ziegenbein and K. H. Hornung (to Chemische Werke Huels AG).

84. **U.S. Pat. 3,406,212** (Oct. 15, 1968), K. O. Christie and A. E. Pavlath (to Stauffer Chemical Co.).

85. **Ger. Pat. 1,133,716** (July 26, 1962), D. Ludsteck, G. Neubauer, H. Pasedach, and M. Seefelder (to Badische Anilin & Soda Fabrik AG).

86. **U.S. Pat. 3,277,143** (Oct. 4, 1966), H. Tilles (to Stauffer Chemical Co.).

87. **U.S. Pat. 3,093,537** (June 11, 1963), H. Tilles.

88. **U.S. Pat. 2,552,591** (May 15, 1951), W. H. C. Rueggeberg, R. K. Frantz, and A. Ginsburg.

89. F. Runge, H. Reinhard, and G. Kьhnhanss, *Chem. Tech. (Berlin)* **8**, 644 (1956).

90. **Ger. Pat. 1,148,073** (May 2, 1963), R. Moroni and E. Dumont (to Collo-Rheincollodium Koeln GmbH).

91. **U.S. Pat. 3,412,072** (Nov. 19, 1968), C. J. Bouboulis and I. Kirschenbaum (to Esso Research & Engineering Co.).

Kenneth L. Dunlap
Bayer Corporation

PHOSPHAMIDON.

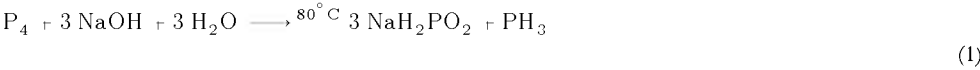
See INSECT CONTROL TECHNOLOGY.

PHOSPHINE AND ITS DERIVATIVES

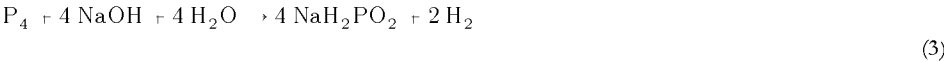
Although phosphine [7803-51-2] was discovered over 200 years ago in 1783 by the French chemist Gingembre, derivatives of this toxic and pyrophoric gas were not manufactured on an industrial scale until the mid- to late 1970s. Commercial production was only possible after the development of practical, economic processes for phosphine manufacture which were patented in 1961 (1) and 1962 (2). This article describes both of these processes briefly but more focus is given to the preparation of a number of novel phosphine derivatives used in a wide variety of important commercial applications, for example, as flame retardants (qv), flotation collectors, biocides, solvent extraction reagents, phase-transfer catalysts, and uv photoinitiators.

Manufacture of Phosphine

Two processes have been used to manufacture gaseous phosphine on a large scale. These are commonly known as the alkaline (1) and acid processes (2). In the alkaline process, an aqueous solution of sodium hydroxide is allowed to react with molten, yellow phosphorus. A long-chain alcohol is used as a dispersant (3). The reaction to produce phosphine may be represented by equation 1.

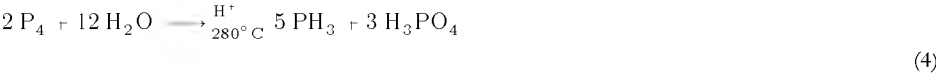


However, hydrogen is formed in two side reactions, ie, by the decomposition of some sodium hypophosphite (eq. 2) and by the direct reaction of phosphorus with sodium hydroxide (eq. 3).



The presence of approximately 60% hydrogen in the gas is a disadvantage because the preparation of many phosphine derivatives involves the high pressure reaction of phosphine with, for example, olefins. However, claims have been made that careful control of the reaction conditions can prevent the production of hydrogen and that 95% pure phosphine can be generated (4,5). Additionally, the stoichiometric yield of phosphine based on phosphorus usage is 25%, but in practice yields of 30% are obtained (3). It is probable, therefore, that equations 1–3 do not represent all of the reaction chemistry. This process, based on the patent, has been operated by Hoechst AG in Germany since the late 1970s.

In the acid process, yellow phosphorus is converted to a mixture of yellow and red by heating to almost 300°C for several hours. The mixture is subsequently treated with steam at 280°C to produce phosphine (eq. 4).



The reaction is acid catalyzed by the by-product phosphoric acid. Only red phosphorus reacts. Unreacted yellow phosphorus is separated and recycled.

The acid process has three advantages over the alkaline process, ie, (1) higher yield of phosphine (60 vs 25%); (2) more pure gas for use in subsequent reactions (95 vs 40%); and (3) by-product phosphoric acid is relatively valuable and can be sold into a number of markets, eg, in the manufacture of fertilizers and flame retardants. There is no ready outlet for the mixture of phosphites produced via the alkaline route and additional processing by oxidative spray drying is needed to produce phosphates for sale (3).

The principal disadvantage of the acid process is the higher capital cost involved; mainly because of more processing steps and the corrosivity of hot, concentrated phosphoric acid which requires a reactor built from dense graphite.

The acid process has been operated since 1970 by Cytec Canada Inc. (Niagara Falls, Canada) and since 1980 by Albright and Wilson Ltd. (Oldbury, England). Many of the details of the process are considered to be proprietary because of its specialized nature. Nippon Chemicals has also been producing phosphine, probably by the acid process, in Japan since the early to mid-1980s. Typical properties of phosphine are given in Table 1.

Table 1. Properties of Phosphine

Property	Value	
appearance and odor	colorless gas with garlic odor	
freezing point, °C	133.8	
boiling point, °C	87.8	
critical temperature, °C		51
critical pressure, MPa ^a		6.485
heat of fusion, kJ mol ^b		1.13
heat of vaporization, kJ mol ^b		14.6
heat of formation, kJ mol ^b		9.59
index of refraction (liquid), n _D [†]		1.317
solubility ^c in water, mL 100 mL H ₂ O		26
heat capacity (liquid), J g		
at 0°C		2.43
at 25°C		2.76
viscosity (gas), mPa(cP)		

at 0°C	0.01
at 100°C	0.014
viscosity (liquid), mPa(=cP)	
at 50C	0.1
at 25°C	0.05
surface tension (liquid), mW m (=dyn cm)	
at 0°C	70
at 25°C	35

^a To convert MPa to psi, multiply by 145.

^b To convert kJ to kcal, divide by 4.184.

^c At 101.3 kPa (1 atm).

Health and Safety Factors

Toxicity. Lethality is the primary hazard of phosphine exposure. Phosphine may be fatal if inhaled, swallowed, or absorbed through skin. All phosphine-related effects seen at sublethal inhalation exposure concentrations are relatively small and completely reversible. The symptoms of sublethal phosphine inhalation exposure include headache, weakness, fatigue, dizziness, and tightness of the chest. Convulsions may be observed prior to death in response to high levels of phosphine inhalation. Some data are given in Table 2.

Table 2. Effect of Phosphine Exposure

Parameter	Concentration, ppmv
TLV–TWA ^c	0.3
odor threshold	1.5–3.0
transient health effects after several hours	5.0–7.0
maximum exposure for 0.5–1 h without serious effects	100–200
lowest lethal concentration (human), 5 min	1000

^a Threshold limit value–time weighted average. Defined as the maximum time weighted average concentration to which a worker may be exposed repeatedly and without adverse effects for a normal 8 h d, 40 h wk period.

The potential mutagenicity of phosphine has been examined by exposing bacteria, tissue culture cells, and animals to phosphine, then examining these organisms for signs of genetic mutation. The majority of these studies have indicated that phosphine is not mutagenic but some of the studies have indicated that phosphine may be a weak mutagen. There is not sufficient evidence to indicate that phosphine is either an animal or human carcinogen.

Safety. The pyrophoric and toxic nature of phosphine requires the adoption of special precautions to ensure safety during manufacture on a commercial scale. Of particular note are the provisions of flame retardant, protective clothing for operating personnel, and strategically located breathing-air stations equipped with in-line respirators. Additionally, the facility operated by Cytec Canada Inc. is equipped with two special phosphine detectors, manufactured by HNU Systems Inc. (Newton, Massachusetts), to protect operators from overexposure. In a cycle which lasts only 12 minutes, each detector automatically collects and analyzes air samples from 10 remote sampling points located in key areas of the plant. The instruments automatically activate audio and visual alarms if the local phosphine concentration exceeds its TLV (0.3 ppm) thus allowing remedial steps to be taken.

Uses

Apart from the manufacture of derivatives, there are only two known uses for phosphine itself, ie, in the preparation of semiconductors and as a fumigant.

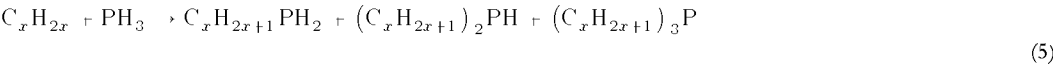
Semiconductors. Phosphine is commonly used in the electronics industry as an *n*-type dopant for silicon semiconductors (6), and to a lesser extent for the preparation of gallium–indium–phosphide devices (7). For these end uses, high purity, electronic-grade phosphine is required; normally >99.999% pure. The main impurities that occur in phosphine manufactured by the acid process are nitrogen [7727-37-9], hydrogen [1333-74-0], arsine [7784-42-1], carbon dioxide [124-38-9], oxygen [7782-44-7], methane [74-82-8], carbon monoxide [630-08-0], and water [7732-42-1]. Phosphine is purified by distillation under pressure to reduce the level of these compounds to <1 ppm by volume. The final product is sold as CYPURE (Cytec Canada Inc.) phosphine.

Fumigants. Phosphine generated *in situ* by the reaction of atmospheric moisture with pelletized calcium, aluminum, or magnesium phosphide is used as a fumigant in, for example, grain silos (8,9). However, this technique suffers from several disadvantages. For example, the rate of phosphine generation and concentration in the atmosphere are dependent on the prevailing moisture content of the atmosphere and the reaction is sufficiently exothermic to cause fire and explosions if free water comes into contact with the pellets (9). A safer, more efficient method has been developed (9) and involves the controlled injection of a mixture containing 2° by weight phosphine in carbon dioxide into the structure to be fumigated. This is used commercially in Australia.

Phosphine Derivatives

Commercial phosphine derivatives are produced either by the acid-catalyzed addition of phosphine to an aldehyde or by free-radical addition to olefins, particularly α-olefins. The reactions usually take place in an autoclave under moderate pressures (<4 MPa (580 psi)) and at temperatures between 60 and 100°C.

In the case of olefins, the reaction generally yields a mixture of primary, secondary, and tertiary phosphines, as follows:

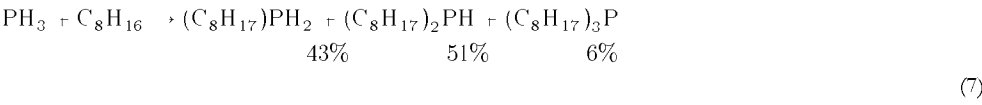


However, the composition of the mixture can be controlled to some extent by the correct choice of olefin and reaction conditions. For example, the production of tertiary phosphines can be maximized by conducting the reaction at relatively low phosphine pressures, ~1.5 MPa (200 psi), and using a 20–30° stoichiometric excess of a straight-chained olefin as in the synthesis of tributylphosphine [988-40-3] by reaction with 1-butene [106-98-9].



In this case, yields >95% of the tertiary phosphine are obtained. Tributylphosphine is readily converted to tetraalkylphophonium salts by reaction with an alkyl halide. These compounds are used commercially as biocides and phase-transfer catalysts.

In contrast, if the olefin is more sterically hindered (branched) and or the reaction is operated at a higher pressure (4 MPa), formation of the primary and secondary phosphines is favored as in the reaction with 2,4,4-trimethyl-1-pentene [107-39-1].



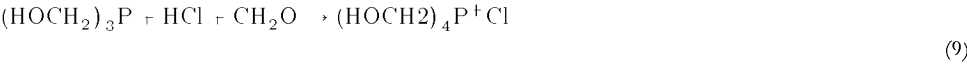
The mixture can be separated by distillation. The primary phosphine is recycled for use in the subsequent autoclave batch, the secondary phosphine is further derivatized to the corresponding phosphinic acid which is widely employed in the industry for the separation of cobalt from nickel by solvent extraction. With even more hindered olefins, such as cyclohexene [110-83-8], the formation of tertiary phosphines is almost nondetectable.

Other typical alkylphosphines that can be prepared through phosphine chemistry are monoisobutylphosphine [4023-52-3], trioctylphosphine [4731-53-7], monocyclohexylphosphine [822-68-4], dicyclohexylphosphine [829-84-5], and triethylphosphine [554-70-1].

Textile Flame Retardants. The first known commercial application for phosphine derivatives was as a durable textile flame retardant for cotton and cotton–polyester blends. The compounds are tetrakis(hydroxymethyl)phosphonium salts (10) which are prepared by the acid-catalyzed addition of phosphine to formaldehyde. The reaction proceeds in two stages. Initially, the intermediate tris(hydroxymethyl)phosphine [2767-80-8] is formed.

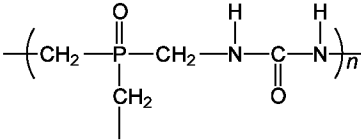


This compound is unstable, particularly at alkaline pH, and decomposes to release hydrogen. It is not isolated but reacts *in situ* with an additional mole of formaldehyde and a mineral acid, for example hydrogen chloride [7647-01-1], to form the phosphonium salt.



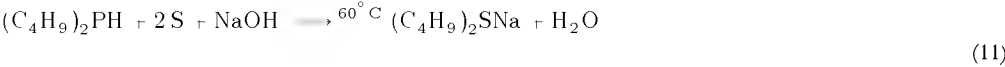
The salt in this case is tetrakis(hydroxymethyl)phosphonium chloride [124-64-1]. The corresponding sulfate salt [55566-30-8] is also produced commercially as are urea-containing formulations of both salts. The latter formulations are actually used to flame retard the textiles (see FLAME RETARDANTS FOR TEXTILES).

After application to the fabric, the compounds are polymerized by reaction with gaseous ammonia (11,12), then oxidized to phosphine oxides by reaction with hydrogen peroxide. The structure of the polymer is shown (13).



This provides a durable finish which, unlike many other flame retardants, can withstand repeated (50–100) launderings without a loss of efficiency. An added advantage is that the feel of the cloth (hand) is little effected. Principal markets are in the treatment of industrial protective clothing, military uniforms, and, in Europe, for furnishings. These products are available from Albright & Wilson Ltd. and Cytec Industries Inc.

Flotation Reagents. Only one sulfide mineral flotation collector is manufactured from phosphine, ie, the sodium salt of bis(2-methylpropyl)phosphinodithioic acid [13360-78-6]. It is available commercially from Cytec Industries Inc. as a 50% aqueous solution and is sold as AEROPHINE 3418A promoter. The compound is synthesized by reaction of 2-methyl-1-propene [115-11-7] with phosphine to form an intermediate dialkylphosphine which is subsequently treated with elemental sulfur [7704-34-9] and sodium hydroxide [1310-73-2] to form the final product (14). The reactions described in equations 10 and 11



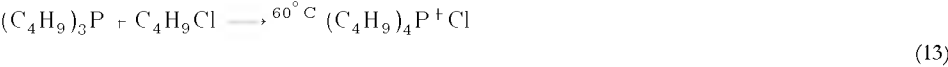
are carried out in an autoclave and a glass-lined kettle, respectively. The primary phosphine formed during the autoclave reaction is removed from the autoclave liquor by distillation and is recycled for use in the next autoclave batch. The tertiary phosphine reacts with sulfur in equation 12 to produce tris(2-methylpropyl)phosphine sulfide [3982-87-4], a solid which is separated from the product using a centrifuge.



AEROPHINE 3418A promoter is widely used in North and South America, Australia, Europe, and Asia for the recovery of copper, lead, and zinc sulfide minerals (see FLOTATION). Advantages in comparison to other collectors (15) are said to be improved selectivity and recoveries in the treatment of complex ores, higher recoveries of associated precious metals, and a stable grade–recovery relationship which is particularly important to the efficient operation of automated circuits. Additionally, AEROPHINE 3418A is stable and, unlike xanthates (qv), does not form hazardous decomposition products such as carbon disulfide. It is also available blended with other collectors to enhance performance characteristics.

Phase-Transfer Catalysts. The use of phase-transfer catalysts to improve kinetics and yields in heterogeneous reactions has been growing rapidly since the 1960s. The five to ten commercial processes in use in 1970 had risen to 550 (16) by 1989. The principal areas of application are in the preparation of polymers, accounting for 50% of catalyst consumption, followed by pharmaceuticals (20%) and agricultural chemicals (10%). Details of the chemistry and applications have been given elsewhere (17) (see CATALYSIS, PHASE TRANSFER). The most common phase-transfer catalysts are quaternary ammonium salts containing either alkyl or mixed alkaryl groups. However, these compounds are being displaced in some applications by the corresponding phosphonium salts mainly because of the enhanced thermal stability of the phosphorus compounds (17). Additionally, the phosphonium salts tend to be more efficient than the nitrogen-based analogues and can promote more rapid reaction kinetics (18).

Phosphonium salts are readily prepared by the reaction of tertiary phosphines with alkyl or benzylic halides, eg, the reaction of tributylphosphine [998-40-3] with 1-chlorobutane [109-69-3] to produce tetrabutylphosphonium chloride [2304-30-5].



Kinetics are slow and many hours are required for a 95% conversion of the reactants. In the case of the subject compound, there is evidence that the reaction is autocatalytic but only when approximately 30% conversion to the product has occurred (19). Reaction kinetics are heavily dependent on the species of halogen in the alkyl halide and decrease in the order I > Br > Cl. Tetrabutylphosphonium chloride exhibits a high solubility in a variety of solvents, for example, ~80% in water, ~70% in 2-propanol, and ~50% in toluene at 25°C. Its analogues show similar properties. One of the latest applications for this phosphonium salt is the manufacture of readily dyeable polyester yarns (20,21).

In addition to tetrabutylphosphonium chloride, typical phosphonium salts that can be produced include tetraoctylphosphonium bromide [23906-97-0], tetrabutylphosphonium acetate [17786-43-5] (monoacetic acid), and tetrabutylphosphonium bromide [3115-68-2]. In most cases, these compounds can be prepared with alternative counterions.

Biocides. Two phosphine derivatives are in commercial use as biocides. These are tetrakis(hydroxymethyl)phosphonium sulfate [55566-30-8] and tributyl(tetradecyl)phosphonium chloride [8741-28-8]. These compounds are sold by Albright and Wilson Ltd. and FMC, respectively. The preparation

of the hydroxymethylphosphonium salt has been discussed (see FLAME RETARDANTS). Synthesis of the tetraalkylphosphonium chloride follows the reaction described in equation 13 except that 1-chlorotetradecane [2425-54-9] is employed in place of 1-chlorobutane.

Various patents (22–24) have been issued claiming the use of tetrakis(hydroxymethyl)phosphonium sulfate in, for example, water treating, pharmaceuticals (qv), and in the oil industry where this compound shows exceptional activity toward the sulfate-reducing bacteria that are a primary cause of hydrogen sulfide formation and consequent problems associated with souring and corrosion (25).

Tributyl(tetradecyl)phosphonium chloride was developed specifically (26,27) as a broad-spectrum biocide for the control of biological fouling in cooling water systems where it is particularly effective (see INDUSTRIAL ANTIMICROBIAL AGENTS).

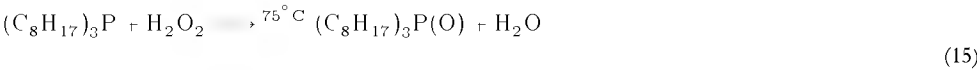
Ultraviolet Photoinitiators. Photoinitiators are used in increasing volume for a multitude of applications. The most important of these are in the formulation of uv-curable inks and in the production of coatings on vinyl flooring, wood, and electronics components (28,29). The most common types of photoinitiators are phenone derivatives, for example, acetophenones and benzophenones (30).

However, Ciba-Geigy has introduced (31,32) a type of phosphine-based photoinitiator. In general, the compound can be described as a bis(acyl)phosphine oxide and is prepared by the reaction of a monoalkylphosphine with a substituted benzoyl chloride (33). The composition of the first commercial product is proprietary.

However, advantages in comparison with conventional photoinitiators, including monoacylphosphine oxides, lie in the ability to prepare thicker coatings that have improved scratch resistance and do not yellow with age. The compound is self-bleaching. Pigmented coatings can also be prepared. This enables formulators, for the first time, to prepare white, uv-cured coatings. Initial areas of application are for furniture coatings and white screen inks.

Solvent Extraction Reagents. Solvent extraction is a solution purification process that is used extensively in the metallurgical and chemical industries. Both inorganic (34,35) and organic (36) solutes are recovered. The large commercial uses of phosphine derivatives in this area involve the separation of cobalt [7440-48-4] from nickel [7440-02-0] and the recovery of acetic acid [61-19-7] and uranium [7440-61-1].

Uranium Recovery From Wet-Process Phosphoric Acid. In the mid- to late 1960s, work at the Oak Ridge National Laboratory (37,38) led to the invention of a process to recover the low concentrations (100–200 mg L) of uranium [7440-61-1], which occur naturally in the wet-process phosphoric acid [7664-38-2] used to make fertilizers (qv). Key to the development of this process was the discovery of the synergic interaction between the bis(2-ethylhexyl) ester of phosphoric acid [298-07-7] (D2EHPA) and trioctylphosphine oxide [78-50-2] (TOPO) in extracting U⁺ (37). D2EHPA is prepared by conventional organophosphorus chemistry and TOPO is readily manufactured by the reaction of phosphine with octene [25377-83-7] to form intermediate trioctylphosphine [4731-53-7] which is subsequently oxidized to TOPO with hydrogen peroxide [7722-84-1] as outlined in equations 14 and 15. TOPO is a white, waxy solid with a melting point of approximately 50°C.



Following further development (38), a two-cycle process has been adopted by industry. In the first concentration cycle, the clarified feed acid containing 100–200 mg L U₃O₈ [1334-59-8] is oxidized, for example, with hydrogen peroxide or sodium chlorate [7775-09-9] to ensure that uranium is in its 6+ valence state; U⁴⁺ is not extracted. Uranium is extracted with a solvent composed of 0.5 *M* D2EHPA and 0.125 *M* TOPO dissolved in an aliphatic hydrocarbon diluent.

Uranium is subsequently stripped reductively from the loaded solvent using a bleed stream of the raffinate acid to which ferrous iron has been added to reduce uranium to its nonextractable, quadravalent state. Raffinate is acid from which uranium has already been extracted. By controlling the organic-to-aqueous volume phase ratios in the extraction and stripping circuits, uranium is concentrated by a factor of approximately 70.

Raffinate acid from the first cycle, containing approximately 7 to 14 g L U₃O₈ is then reoxidized and re-extracted in the second, purification cycle using a solvent containing 0.3 *M* D2EHPA and 0.075 *M* TOPO. The loaded solvent is washed with iron-free acid to remove iron and then with water to remove extracted and entrained acid. The solvent is stripped with ammonium carbonate [506-87-6] to yield ammonium uranyl tricarbonate [18077-77-5] which is subsequently calcined to U₃O₈ (yellow cake). The stripped solvent is regenerated with mineral acid before recycling (39).

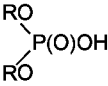
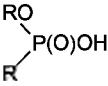
Beginning in approximately 1975, both IMC and Freeport Minerals operated large uranium recovery plants in the United States using this technology. Several plants continue to run but a number have been closed because of the depressed uranium prices that resulted when uranium from the former Soviet Union flooded Western markets. A relatively small plant is operated by Prayon in Belgium (40). TOPO is available from Cytec Industries Inc. as CYANEX 921 extractant. D2EHPA is available from Albright & Wilson Ltd. and is also sold by Daihachi as DP-8R.

Acetic Acid Recovery. Sulfite wood pulping operations produce dilute, aqueous effluents containing 10–20 g L acetic acid. In some cases, 2-furancarboxaldehyde [98-01-1], more commonly known as furfural, can also be present at lower concentrations (~1 g L) (41). Lenzing (Austria) recovers both of these by-products by solvent extraction with TOPO. Although few data concerning the plant have been published (41,42), it is known (43) that the solvent is 30% TOPO in undecane [1120-21-4]. The extraction column is operated at 50°C and the aqueous-to-organic volume phase ratio (A/O) is 1. The loaded solvent is distilled to strip the extracted species, first to remove most of the water for recycle, then to strip an azeotrope of water, acetic acid, and furural. The azeotrope is further distilled to yield pure acetic acid and furfural. Both compounds are sold. The plant has been operating successfully since 1983 and supplies approximately 50% of Austria's demand for food-grade acetic acid.

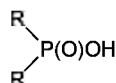
In a similar application, Cape Industries has announced its intention to commission a solvent extraction plant to recover acetic acid from an effluent generated at its dimethyl terephthalate [120-61-6] facility (Wilmington, North Carolina) (44,45). The plant was commissioned in February 1995. In this case, the solvent will be CYANEX 923 extractant [100786-00-3]. CYANEX 923 is also a phosphine oxide, but unlike TOPO is a liquid and can be used without a diluent (46,47). This has the benefit of reducing plant size, capital, and operating costs.

Cobalt–Nickel Separation. The bis(2,4,4-trimethylpentyl)phosphinic acid [83411-71-6] became commercially available during the early 1980s (48,49). It is sold by Cytec Industries Inc. as CYANEX 272 extractant and was developed specifically to selectively extract cobalt from weakly acidic, nickeliferous solutions. It is a member of one of three groups of organophosphorus extractants that have been examined for cobalt–nickel separation. These are derivatives of phosphoric (50), phosphonic (51), and phosphinic (52) acids. CYANEX 272 has two significant advantages over its competing reagents. The first is superior cobalt–nickel selectivity, as illustrated by the results of some batch equilibrium tests shown (52) in Table 3.

Table 3. Organophosphorus Extractants for Co–Ni Separation^a

Extractant ^b type	Structure	Commercial homologue, R	Co–Ni ^c separation factor
phosphoric acid (D2EHPA)		2-ethylhexyl	14
phosphonic acid (PC-88A)		2-ethylhexyl	280

phosphinic acid (CYANEX 272)



2,4,4-trimethylpentyl

7000

^a Conditions: temperature 25°C; equilibrium pH 4; A O 1.^b 0.1 M extractant in MSB 210.

^c Each metal ion concentration $2.5 \times 10^{-2} M$.

The benefits of high selectivity lie in the ability to produce high purity cobalt in a limited number of stages. This minimizes capital and operating costs. It is particularly important when the solution in question contains low concentrations of cobalt. For example, solutions derived from laterite deposits may only contain 0.5–2 g L Co but 90–100 g L Ni.

The second principal advantage is that CYANEX 272 is the only one of the three above-mentioned compounds that extracts cobalt in preference to calcium (52). This property can minimize or eliminate the solvent losses that are associated with calcium extraction and the subsequent precipitation of gypsum cruds in the scrubbing or stripping circuits. This is illustrated in Figure 1 where calcium extraction is shown as a function of pH for the three subject reagents.

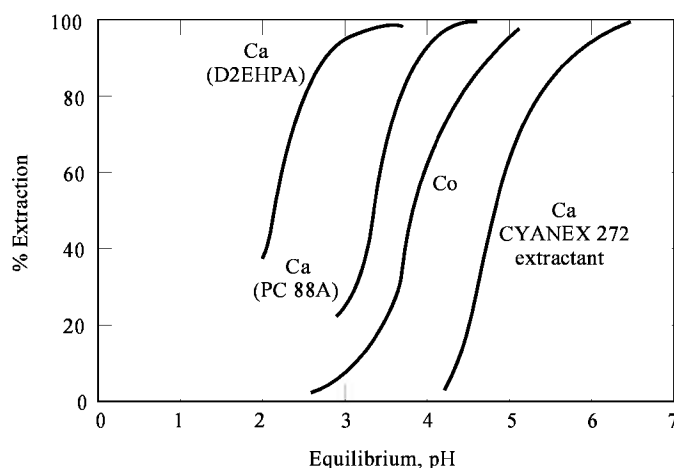


Fig. 1. Cobalt-calcium selectivity with organophosphorus extractants. Conditions: solvent 0.6 M extractant in Kermac 470B; aqueous 0.015 M metal ion as sulfate; temperature 50°C; and A/O = 1.

The first commercial plant to use CYANEX 272 became operational in 1985. An additional three plants were constructed between 1985 and 1989. Of the four, one is in South America and three in Europe. An additional three plants have been built; two in Europe (1994) and one in North America (1995). Approximately 50% of the Western world's cobalt is processed using CYANEX 272. Both high purity salts and electrolytic cobalt metal are recovered from solutions ranging in composition from 30 g/L each of cobalt and nickel to 0.2 g/L Co, 95 g/L Ni. Operating companies usually regard use of CYANEX 272 as confidential for competitive reasons and identities cannot be disclosed. CYANEX 272 is being evaluated on the pilot-plant scale in many additional projects involving the recovery of cobalt and other metals.

BIBLIOGRAPHY

1. U.S. Pat. **2,977,122** (Mar. 28, 1961), R. W. Cummins (to Food Machinery and Chemical Corp.).
2. Fr. **1,352,605** (Feb. 14, 1964), (to Albright & Wilson Ltd. and Hooker Chemical Corp.).
3. N. Weferling, "Phosphine Based Organophosphorus Products and Their Applications," paper presented at *Chemspec Europe '88*, Frankfurt, Germany, Mar. 22-23, 1988.
4. Ger. Offen. DE **2,632,316** (Jan. 19, 1978), J. Stenzel, G. Heymer, and C. May (to Hoechst AG).
5. N. Weferling, *Phos. Sulf.* **30**, 641 (1987).
6. S. Lian and co-workers, *J. Vac. Sci. Technol.* **11**(6), 2914 (1993).
7. J. F. Carlin and co-workers, *Mater. Sci. Eng.* **B21**(2-3), 293 (1993).
8. *Chem. Weeks*, **134**(14), 14 (Apr. 1984).
9. Brit. Pat. Appl. **2,177,004A** (Jan. 14, 1987), S. Latif and R. F. Ryan (to The BOC Group PLC).
10. U.S. Pat. **3,888,779** (June 10, 1975), C. T. Hsiang (to Cytec Technology Corp.).
11. Eur. Pat. Appl. EP **294234** A2 (July 12, 1988), G. W. Smith (to Albright and Wilson, Ltd.).
12. Can. Pat. Appl. Ca **2048402** AA (Feb. 11, 1992), R. Cole (to Albright and Wilson Ltd.).
13. W. A. Reeves and R. M. Perkins, *Colourage*, (18), 1-7 (1971).
14. Brit. Pat. Appl. **2,068,381A** (Aug. 12, 1981), A. J. Robertson and T. Ozog (to Cytec Technology Corp.).
15. P. A. Mingione, "Use of AEROPHINE® 3418A Promoter for Sulphide Minerals Flotation," *Proceedings of the 22nd Annual Meeting of the Canadian Mineral Processors*, Ottawa, 1990.
16. *Phase Transfer Catalysis in Industry*, PTC Interface, Inc., Marietta, Ga.
17. C. M. Starks and C. Liotta, *Phase Transfer Catalysis*, Academic Press, Inc., New York, 1978.
18. C. M. Starks, "Selecting a Phase Transfer Catalyst," *Chemtech* (Feb. 1980).
19. A. J. Robertson, private communication, Cytec Canada Inc., Phosphine Technical Centre, Jan. 1994.
20. Jpn. Kokai Tokyo Koho JP **03241024** (Jan. 28, 1991), M. Yanagihara, K. Kawakami, and H. Nagai (to Taijin Ltd.).
21. Eur. Pat. Appl. EP **280028A2** (Aug. 31, 1998), T. Suzuki and co-workers (to Teijin Ltd.).
22. U.S. Pat. **4,673,509** (June 16, 1987), K. P. Davis and R. E. Talbot (to Albright and Wilson Ltd.).
23. Eur. Pat. Appl. EP **275207** A2 (July 20, 1988), R. E. Talbot and co-workers (to Albright and Wilson Ltd.).
24. U.S. Pat. **4,775,407** (Oct. 4, 1988), K. G. Cooper and R. E. Talbot (to Albright and Wilson Ltd.).
25. *Toxicide PS 72A Product Brochure*, Albright and Wilson Biocides, Oldbury, Worley, West Midlands, U.K.
26. Eur. Pat. Appl. EP **0066544** (May 30, 1981), R. Grade and B. M. Thomas (to Ciba-Geigy AG).
27. Can. Pat. Appl. CA **2082994** AA (May 28, 1993), W. Wehner and R. Grade (to FMC Corp. (U.K.) Ltd.).

28. K. K. Dietliker and P. Oldring, "Chemistry and Technology of UV & EB Formulations for Coatings, Inks & Paints," Vol. 3, *Photoinitiators for Free Radical and Cationic Polymerization*, Sholium International, 1991.

29. K. Lawson, "UV EB Curing in North America—1994," *RadTech Rep.* **21** (Mar. Apr. 1994).

30. N. S. Allen, *Photopolymerization and Photoimaging Science and Technology*, Elsevier Applied Science, London, 1989.

31. Eur. Pat. Appl. **0184095** A2 (Nov. 27, 1984), K. Ellrich and C. Herzig (to Espe Fabrik Pharmazeutischer Preparate GmbH).

32. K. Dietliker and co-workers, "Novel High Performance Bisacylphosphine Oxide (BAPO) Photoinitiators," paper presented at *RadTech'94, Orlando, Florida*, May 1–5, 1994.

33. Ger. Offen., DE **4,231,579** (Mar. 25, 1994) D. G. Leppard, M. Koehler, and L. Misev (to Ciba-Geigy AG).

34. G. M. Ritcey and A. W. Ashbrook, *Solvent Extraction—Principles and Applications to Process Metallurgy*, Part I and II, Elsevier, Amsterdam, the Netherlands, 1979.

35. T. C. Lo, M. H. I. Baird, and C. Hanson, *Handbook of Solvent Extraction*, John Wiley and Sons, Inc., New York, 1983.

36. F. J. Hurst, D. J. Crouse, and K. B. Brown, *Solvent Extraction of Uranium from Wet Process Phosphoric Acid*, ORNL-TM-2522, U.S. Atomic Energy Commission, Washington, D.C., 1969.

37. F. J. Hurst, D. J. Crouse, and K. B. Brown, *Ind. Eng. Chem., Process Des. Develop* **13**, 286–291 (1974).

38. P. D. Mollere, *DECHMA* **2**, 49 (1986).

39. U.S. Pat. **4,105,74** (Aug. 8, 1978), T. K. Wiewioroski and W. L. Thornsberry (to Freeport Minerals Co.).

40. W. W. Berry, *Chem. Eng. Prog.* **77**(2), 76–82 (1981).

41. Eur. Pat. Appl. EP **36406** A1 (Sept. 23, 1981), W. Kanzler and J. Schedler (to Vereinigte Edelstahlwerke AG).

42. Eur. Pat. Appl. EP **38317** (Oct. 21, 1981), W. Kanzler and J. Schedler (to Vereinigte Edelstahlwerke AG).

43. E. K. Watson and W. A. Rickelton, *Solv. Extr. Ion Exch.* **10**(5), 879 (1992).

44. *Chem. Week*, **152**(10), 40 (Mar. 17, 1993).

45. *Oil Gas J.* **91**(15), 35 (Apr. 12, 1993).

46. Technical brochure, *CYANEX®923 Extractant*, SPT-032a, Cytec Industries, Inc., West Paterson, N.J., 1987.

47. U.S. Pat. **4,909,939** (Mar. 20, 1991), W. A. Rickelton and A. J. Robertson (to Cytec Technology Corp.).

48. U.S. Pat. **4,353,883** (Oct. 12, 1982), W. A. Rickelton, A. J. Robertson, and D. R. Burley (to Cytec Technology Corp.).

49. U.S. Pat. **4,374,780** (Feb. 22, 1983), A. J. Robertson (to Cytec Technology Corp.).

50. G. M. Ritcey, A. W. Ashbrook, and B. H. Lucas, *CIM Bull.* **68**, 111–123 (1975).

51. J. S. Preston, *J. S. Afr. Inst. Min. Metall.* **83**, 126–132 (1983).

52. W. A. Rickelton, D. S. Flett, and D. W. West, *Solv. Extr. Ion Exch.* **2**(6), 815–838 (1984).

William A. Rickelton
Cytec Canada Inc.

PHOSPHORIC ACIDS AND PHOSPHATES

Phosphoric acids,
Phosphates,
Economic aspects,
Safety and environmental considerations,

Phosphorus (qv), in the form of phosphate, is an element essential to all life and the use of natural phosphatic fertilizers (qv) such as bones, fish, and guano predates recorded history. The birth of the modern phosphate industry occurred in the mid-nineteenth century through the manufacture of phosphatic fertilizer. Sulfuric acid was used for decomposing bones and mineral phosphates to render the phosphate value in a more concentrated and available form. As the production of fertilizers grew more economical and higher purity phosphoric derived from elemental phosphorus became available, so did the volume and diversity of industrial applications of phosphoric acid and the phosphates. A few of the historically notable nonfertilizer developments include the use of monocalcium phosphate [7758-23-8] (MCP) in baking powder in 1886 (see BAKERY PROCESSES AND LEAVENING AGENTS) and the commercial production of elemental phosphorus in 1890.

Most commercial applications of phosphorus-containing materials are based on phosphoric acid and phosphate salts. A multitude of applications exist, including use in detergents (see DETERGENCY), animal feed supplements (see FEEDS AND FEED ADDITIVES), dentifrices (qv), fertilizers (qv), metal treating (see METAL SURFACE TREATMENTS), water softening, leavening agents, and fire retardants (see FLAME RETARDANTS). On a weight basis, fertilizers remain the single largest application. The manufacture and uses of phosphoric acid and phosphate salts for most applications are usually dependent on the purity of the acid or salt. The chemistry and commercial technology of technical- and food-grade phosphoric acid and the inorganic phosphates is discussed herein. Wet-process phosphoric acid and fertilizer phosphates are discussed elsewhere (see FERTILIZERS).

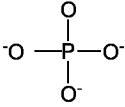
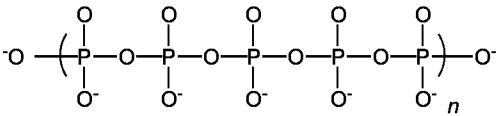
Phosphoric acids and the phosphates may be defined as derivatives of phosphorus oxides where the phosphorus atom is in the +5 oxidation state. These are compounds formed in the $M_2O-P_2O_5$ system, where M represents one cation equivalent, eg, H^+ , Na^+ , $0.5\ Ca^{2+}$, etc. The molecular formula of the phosphorus(V) oxide [1314-56-3] is actually P_4O_{10} , but this oxide is commonly referred to in terms of its empirical formula, P_2O_5 . Structurally, four phosphorus–oxygen (P–O) linkages are arranged in an approximate tetrahedral configuration about the phosphorus atom in the phosphate anion. Compounds containing discrete, monomeric PO_4^{3-} ions are known as orthophosphates or simply as phosphates.

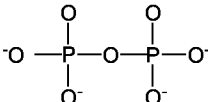
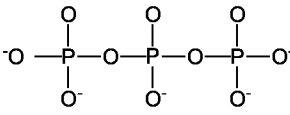
Orthophosphoric acid [7664-38-2], H_3PO_4 , can be considered the building block from which other phosphoric acids and the phosphate salts are derived through the basic reactions of polymerization and/or neutralization. Polymerization occurs via dehydration, hence the polymers are known generally as condensed phosphates. Oxygen atoms shared between PO_4 tetrahedra lead to polymers built on a covalent backbone of P–O–P linkages. Phosphates containing even one P–O–P linkage are also included as condensed phosphates because these are considered to be molecularly dehydrated orthophosphate monomers. In a more specific consideration of the structure, linear P–O–P chains are termed polyphosphates, cyclic rings are known as metaphosphates, and branched polymeric materials and cage anions containing at least one triply linked phosphate monomeric unit are known as ultraphosphates. Structural studies of crystalline condensed phosphates have shown that each PO_4 group remains approximately tetrahedral. The O–P–O bond angles are 95–125°; P–O–P angles vary between 120–180°.

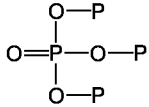
Traditionally, phosphates have been represented as stoichiometric combinations of oxides. It is common practice to speak of the phosphates in terms of oxide ratios; for example, phosphoric acid [7664-38-2], H_3PO_4 , as $1.5\ H_2O\cdot0.5P_2O_5$; disodium phosphate [7558-79-4], Na_2HPO_4 , as $Na_2O\cdot0.5H_2O\cdot0.5P_2O_5$; and sodium triphosphate [7601-54-9], $Na_5P_3O_{10}$, as $2.5\ Na_2O\cdot1.5P_2O_5$.

The oxide ratio, R, is defined as the mole ratio of M_2O to P_2O_5 , where M represents one cation equivalent, metal or hydrogen, including the H_2O of composition, but not water of hydration. The oxide ratio determines the structural type for a pure phosphate compound. The structural designations, oxide ratios, and general formulas for the phosphate anions are shown in Table 1. If the oxide mole ratio is three, the substance is an orthophosphate, commonly known simply as phosphate. Linear polymers, known generally as polyphosphates, have the general formula, $M_{n+2}P_nO_{3n+1}$. The R ranges from 1 to 2, depending on the chain length. An additional term, oligophosphates, is sometimes applied to polyphosphate chains of short or intermediate length. Common short-chain anions include the pyrophosphate, also called diphosphate, and tripolyphosphate (triphosphate). When the chain length, *n*, becomes large, the polyphosphate composition approaches that of the metaphosphate ring structures. Hence, it had been common practice in the phosphorus literature to designate polyphosphates having long chains as metaphosphates. These long-chain polyphosphates are sometimes also called pseudometaphosphates. The term metaphosphate should, however, be reserved for cyclic structures having an R of one and the exact composition, $(MPO_3)_n$, if a ratio of exact unity corresponds to the cyclic metaphosphates. When $0 < R < 1$, the substance contains branching phosphate groups and is called an ultraphosphate.

Table 1. Classification of Phosphate Anions

Designation	Oxide ratio, R	General formula	Structure comment
phosphate + metal oxide orthophosphate	>3 3	PO_4^{3-}	mixtures; includes double salts and solid solutions 
polyphosphates, <i>n</i> = 2, 3, 4, ...	1–2	$P_nO_{3n+2}^{n+2-}$	

pyrophosphate	2	$P_2O_7^{4-}$	
tripolyphosphate	1.67	$P_3O_{10}^{5-}$	
metaphosphates $n = 3, 4, 5, \dots$	1	$ca (PO_3)_n$	cyclic
very large n	~ 1		very long chain
ultraphosphates, $0 < x < 1$	0–1	$P_2O_5x(O^{2-})$	
phosphorus pentoxide	0	$(P_2O_5)_n$	cross-linked chains and or rings; P_4O_{10} or continuous structures



some

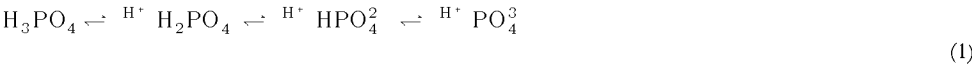
The ultraphosphates are situated between P_4O_{10} and the metaphosphates. These comparatively little-known, highly cross-linked polymers contain at least some of the phosphorus atoms as triply connected branching points. This structural feature is quite unstable toward hydrolysis. Ultraphosphates undergo rapid decomposition upon dissolution. In amorphous ultraphosphates, the cross-linking is presumably scattered randomly throughout the structural matrix; in contrast, crystalline ultraphosphates have a regular pattern.

Many of the terms describing phosphates are used loosely and interchangeably in the technical and trade literature. Herein the terminology most generally used in industry is employed, although cross reference to the most recent *Chemical Abstracts*-approved nomenclature is also given. However, the older naming system is favored. In addition, the phrase metaphosphate composition is used where long-chain structure is uncertain. The term metaphosphate is reserved for true ring systems, a convention that *Chemical Abstracts* follows.

Phosphoric Acids

ORTHOPHOSPHORIC ACID

Properties. Phosphoric acid is a tribasic acid, in which the first hydrogen ion is strongly ionizing, the second moderately weak, and the third very weak.



The titration curve of phosphoric acid in the presence of sodium hydroxide is shown in Figure 1. Three steps, corresponding to consecutive replacement of the three acidic hydrogens, and two inflection points, near pH 4.5 and 9.0, are evident. Dissociation constants are $K_1 = 7.1 \times 10^{-3}$; $K_2 = 6.3 \times 10^{-8}$; and $K_3 = 4.4 \times 10^{-13}$. Both acidic and basic salts can be formed from phosphoric acid, and mixtures of mono- and dibasic salts of H_3PO_4 act as buffers near pH 7 (see HYDROGEN ION ACTIVITY).

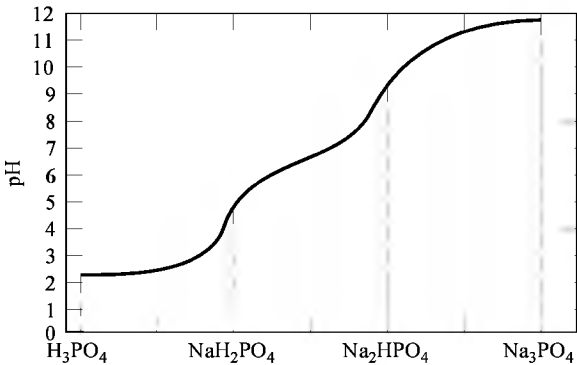


Fig. 1. Titration curve of orthophosphoric acid in the presence of sodium hydroxide.

Phosphoric acid, aside from its acidic behavior, is relatively unreactive at room temperature. It is sometimes substituted for sulfuric acid because of its lack of oxidizing properties (see SULFURIC ACID AND SULFUR TRIOXIDE). The reduction of phosphoric acid by strong reducing agents, eg, H_2 or C, does not occur to any measurable degree below 350–400°C. At higher temperatures, the acid reacts with most metals and their oxides. Phosphoric acid is stronger than acetic, oxalic, silicic, and boric acids, but weaker than sulfuric, nitric, hydrochloric, and chromic acids.

A solid–liquid phase diagram of the phosphoric acid–water system is given in Figure 2 (1). The solid, crystalline phases in the phosphate region of the system are the anhydrous acid, H_3PO_4 , and the hemihydrate [16271-20-8], $H_3PO_4 \cdot 0.5H_2O$. Pure, 100% phosphoric acid is a white, monoclinic crystalline solid that melts at 42.35°C; the hemihydrate has a melting point of 29.25°C. When anhydrous phosphoric acid is melted, reorganization occurs in the liquid phase according to the equilibrium



If the anhydrous acid is maintained in a molten state, the freezing point gradually falls to an equilibrium value of 34.6°C after several weeks; this corresponds to the presence of approximately 6 mol% pyrophosphoric acid [2466-09-3], $H_4P_2O_7$, also known as diphosphoric acid. Thus, sample history affects the melting point of phosphoric acid, which may explain the variation in values reported in the literature. Pure phosphoric acid supercools readily and typically can be stored without crystallization for long periods of time, at 10–20°C below the melting point. When cooled in liquid nitrogen or dry ice, phosphoric acid solidifies to a glass that crystallizes to H_3PO_4 -III upon warming to $-54^\circ C$. This, in turn, is converted reversibly at $-6^\circ C$ to H_3PO_4 -I, which is stable at room temperature; H_3PO_4 -II crystallizes from the melt between 8–15°C if careful temperature control is maintained.

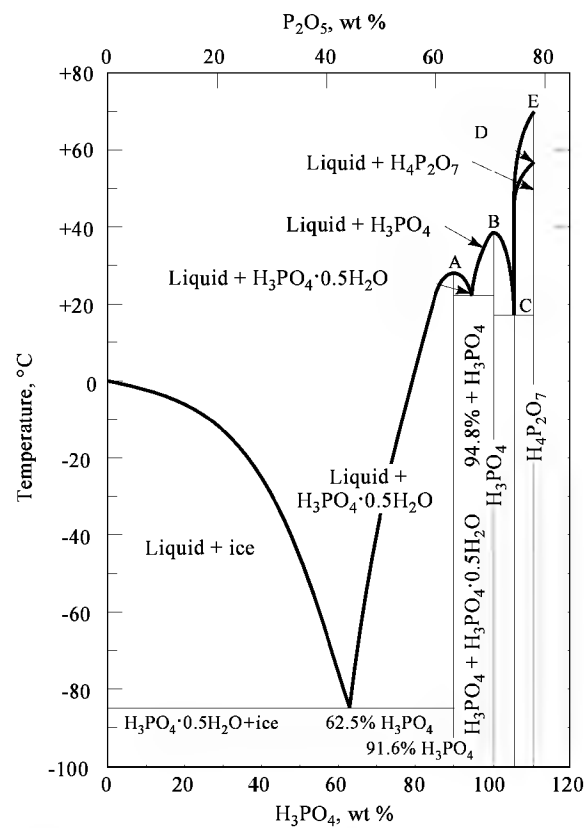


Fig. 2. Solid–liquid phase diagram of the H₂O–P₂O₅ system. The peaks A–E correspond to temperatures of 29.3, 38.85, 16.0, 54.3, and 71.5°C, respectively.

Extensive hydrogen bonding takes place in phosphoric acid solutions. In concentrated (86° H₃PO₄) solutions, as well as in the crystal structures of the anhydrous acid and the hemihydrate, the tetrahedral H₃PO₄ groups are linked by hydrogen bonding. At lower (75° H₃PO₄) concentrations, the tetrahedra are hydrogen-bonded to the water lattice. Physical properties of phosphoric acid solutions of various concentrations are listed in Table 2; the vapor pressure of aqueous H₃PO₄ solutions at various temperatures is given in Table 3.

Table 2. Physical Properties of Aqueous Solutions of Phosphoric Acid

Concentration, wt %		Density at 25°C, g cm ³	Boiling point, °C	Freezing point, °C	Viscosity, mPas(cP)		
H ₃ PO ₄	P ₂ O ₅				20°C	60°C	100°C
0	0	0.997	100.0	0	1.0	0.48	0.30
5	3.62	1.025	100.1	0.8	1.1	0.54	0.33
10	7.24	1.053	100.2	2.1	1.2	0.61	0.38
20	14.49	1.113	100.8	6.0	1.6	0.78	0.48
30	21.73	1.182	101.8	−11.8	2.2	1.0	0.62
50	36.22	1.333	108	−44.0	4.3	1.8	1.1
75	54.32	1.573	135	17.5	15	4.8	2.4
85	61.57	1.685	158	21.1	28	8.1	3.8
100	72.43	1.864	261	42.35	140	25	9.2
105	76.10	1.925	>300	16.0	600	70	1.9
115	83.29	2.044	>500			1500	250

Table 3. Vapor Pressure of Phosphoric Acid Solutions, kPa^a

Concentration, wt %, H ₃ PO ₄	Temperature, °C							
	20°C	30°C	40°C	60°C	80°C	100°C	110°C	140°C
0	2.35	4.24	7.37	20.0	47.3	101.3	143.3	
5	2.33	4.20	7.27	19.6	46.9	100.7	142.4	
10	2.31	4.13	7.23	19.5	46.7	100.4	142.1	
20	2.27	4.00	7.07	18.8	45.5	98.0	138.7	
30	2.17	3.85	6.73	18.1	43.6	94.0	120.8	
50	1.73	3.08	5.37	14.4	34.3	76.7	108.5	
75	0.75	1.33	2.33	6.27	14.8	32.0	45.3	119.3
85	0.29	0.53	0.93	2.63	6.51	14.8	21.3	59.3
100	0.004	0.008	0.016	0.057	0.177	0.487	0.773	2.71

^a To convert kPa to mm Hg, multiply by 7.50.

Manufacture. Phosphoric acid, H₃PO₄, is the second largest volume mineral acid produced; sulfuric acid is the first. The greatest consumption of phosphoric acid is in the manufacture of phosphate salts, as opposed to direct use as acid. Markets are differentiated according to the purity of the acid. Phosphoric acid is produced commercially by either the wet process or the thermal (furnace) process. Thermal acid, manufactured from elemental

phosphorus, is more expensive and considerably purer than wet-process acid. Thermal acid is produced in much smaller quantities than wet-process acid, which is produced directly from phosphatic ores, and is characterized by relatively high production volume, low cost, and low purity. Wet-process acid is used primarily in the production of fertilizers and animal feed supplements. Wet-acid purification has been practiced extensively outside the United States and Canada, but was introduced to the United States in an ongoing commercial scale by Purified Acid Partners (a partnership between Texasgulf Inc. and Albright & Wilson Americas Inc.) and Olin Corp. in 1990, and by Rhone-Poulenc Basic Chemicals Co. in 1991. A comparative analysis of typical wet-process acid, purified wet acid, and thermal acid is given in Table 4. Both thermal and purified wet-process phosphoric acid (WPA) are used almost exclusively in various technical and food applications where fertilizer-grade wet acid is not suitable.

Table 4. Typical Analysis of Phosphoric Acids, wt %

Assay	Wet-process acid		Thermal acid
	Merchant-grade ^a	Technical-grade ^{b,c}	Technical-grade ^d
P ₂ O ₅	53.1	57	54.32
CaO	0.06		0.001
F	0.8	0.02	<0.0001
Al ₂ O ₃	1.7		0.0003
Fe ₂ O ₃	1.23	0.004	0.0004
MgO	0.58		0.0002
K ₂ O	0.01		0.0007
Na ₂ O	0.12		0.0025
SiO ₂	0.07		0.0015
SO ₄	2.2	0.04	<0.002

^a Tennessee Valley Authority.

^b Merchant-grade acid is purified to yield technical-grade.

^c Soci  t   Chimique Prayon-Rupel.

^d Monsanto Co.

Thermal Process. In the manufacture of phosphoric acid from elemental phosphorus, white (yellow) phosphorus is burned in excess air, the resulting phosphorus pentoxide is hydrated, heats of combustion and hydration are removed, and the phosphoric acid mist collected. Within limits, the concentration of the product acid is controlled by the quantity of water added and the cooling capabilities. Various process schemes deal with the problems of high combustion-zone temperatures, the reactivity of hot phosphorus pentoxide, the corrosive nature of hot phosphoric acid, and the difficulty of collecting fine phosphoric acid mist. The principal process types (Fig. 3) include the wetted-wall, water-cooled, or air-cooled combustion chamber, depending on the method used to protect the combustion chamber wall.

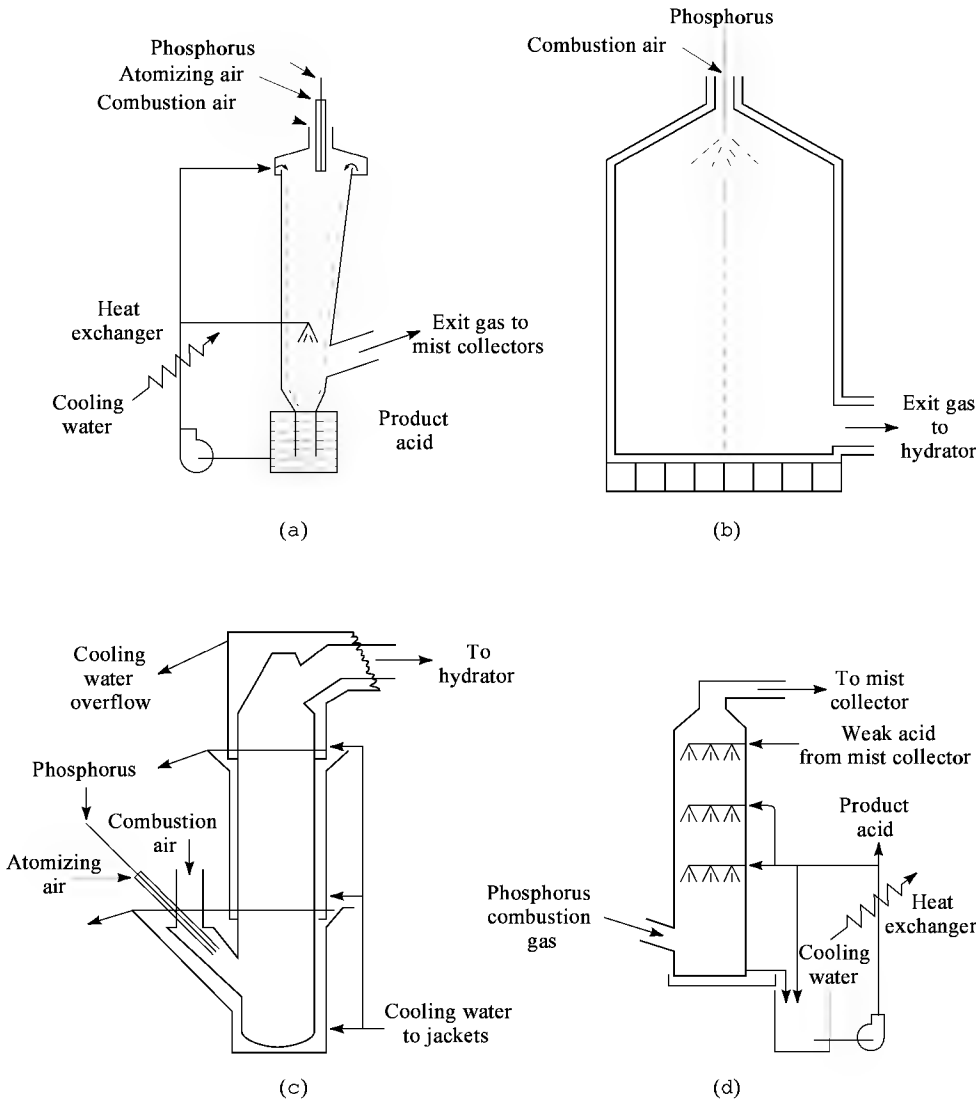


Fig. 3. Thermal phosphoric acid processes: (a) wetted-wall combustion chamber; (b) air-cooled combustion chamber; (c) water-cooled combustion chamber; and (d) hydrator-absorber.

In wetted-wall units, the walls of a tall circular, slightly tapered combustion chamber are protected by a high volume curtain of cooled acid flowing down inside the wall. Phosphorus is atomized by compressed air or steam into the top of the chamber and burned in additional combustion air supplied by a forced or induced draft fan. Wetted-wall plants use 25–50% excess combustion air to reduce the tail-gas volume, resulting in flame temperatures in excess of 2000°C. The combustion chamber may be refractory lined or made of stainless steel. Acid sprays at the bottom of the chamber or in a subsequent, separate spraying chamber complete the hydration of phosphorus pentoxide. The sprays also cool the gas stream to below 100°C, thereby minimizing corrosion to the mist-collecting equipment (typically type 316 stainless steel).

When all acid is to be converted into sodium phosphate salt, a variation of the wetted-wall acid plant may be used (2). In this case, a relatively noncorrosive, neutral sodium phosphate solution is circulated in lieu of phosphoric acid. By operating this system at the solution boiling point, heats of combustion, hydration, and neutralization are removed via evaporative cooling, resulting in a more energy-efficient process. Phosphorus pentoxide absorption is rapid and over 95% is collected by the circulating stream. Alkali and makeup water are added through a tail-gas scrubber as dilute caustic soda, soda ash, or sodium sesquicarbonate solution adjusted to maintain the system material balance. They must have efficient phosphorus atomization and good control of combustion air and phosphorus feed rates to prevent formation of lower oxides of phosphorus.

In a development pioneered by the Tennessee Valley Authority (TVA), the phosphorus combustion was separated from the wetted-wall hydrator–absorber vessel in order to allow more time for complete combustion and increase the area available for heat transfer (3). Both vessels are constructed of extra low carbon type 316 stainless steel and are cooled by an external water jacket. A solid layer of combustion products deposits on the inner surfaces of the combustion chamber and serves as a self-repairing protective coating. This layer provides resistance to heat transfer and consequently the stainless steel wall remains near the jacket water temperature. Although the composition of the deposit varies with the amount of water introduced with the phosphorus and combustion air, it is known to be a mixture of different solid-phase forms (O and O') of phosphorus pentoxide embedded in a glassy matrix (see PHOSPHORUS COMPOUNDS). These plants are quiet and clean, have high capacity, and are capable of producing both orthophosphoric acid and polyphosphoric acid [8017-16-1] at any strength up to about 117% H₃PO₄ (85% P₂O₅).

Air-cooled acid plants are characterized by a large refractory-lined combustion chamber from which waste heat is removed by radiation and convection. The combustion chamber is constructed of graphite or of carbon steel lined with a single layer of high alumina refractory brick. Refractory units operate at cooler temperatures because of the poorer heat transfer properties of brick compared to graphite. Corrosion of the carbon steel is, surprisingly, not a serious problem as long as the combustion gas stream and refractory stay well-above the dew point of the azeotropic (92% P₂O₅) phosphoric acid. Air-cooled plants normally operate with about 1–200% excess combustion air to reduce the flame temperature (1000–1700°C) and carry waste heat to the hydrator–absorber, where it is removed by evaporation of water.

Hot combustion gases are quenched and saturated with water in a spray chamber called a hydrator. An absorber bed of carbon or graphite rings may be mounted above the hydrator in the same structure to obtain more complete absorption of P₄O₁₀ and to assure that the gas stream is cooled to about 100°C. Weak acid from mist collection is sprayed on the absorber bed, and product acid at 75–85% H₃PO₄ leaves the hydrator through a heat exchanger.

Burning phosphorus produces a persistent white cloud of phosphorus pentoxide and phosphoric acid droplets of such high obscuring power that this cloud is used as a standard military screening smoke (see CHEMICALS IN WAR). The opacity and persistence of the smoke results from the whiteness and small size of the particles, most of which are <3 μm in diameter and subject to appreciable Brownian motion. These particles are difficult to collect but an essentially haze-free stack effluent may be obtained to meet air quality regulations, which, for practical purposes, requires an effluent containing <ca 25 milligrams of P₂O₅ per dry standard cubic meter of stack gas.

Because P₂O₅ absorbs more rapidly in strong acid than in water, it is initially hydrated and absorbed in hot gas stream by direct contact with relatively strong acid. This is often followed by successive stages of scrubbing with progressively more dilute acid and, finally, with incoming makeup water. Many of the submicrometer-size droplets are induced to grow large enough for collection by conventional scrubbing techniques. Although this technique collects over 99% of the P₂O₅, it is usually inadequate to meet plume opacity regulations.

Electrostatic precipitators in phosphoric acid service are expensive to maintain and have been largely replaced by newer devices. High energy scrubbers such as the Venturi were used widely at one time, but are generally ineffective for phosphoric acid mist droplets under ca 1 μm. Neither precipitators nor Venturi scrubbers, as single-stage collectors, can reduce the P₂O₅ content of the stack to an acceptable level. Fine-mesh glass fiber or metal-mesh panels can be effective as final-stage collectors when operated at high pressure drop, but usually require a secondary coarse mesh pad to collect reentrained larger particles from the back side of the primary panel. A high efficiency fiber-bed mist eliminator, developed by Monsanto (4), is effective in eliminating mist droplets of <0.3 μm in diameter. By careful choice of fiber diameter, packing density, and bed depth, the gas flow through the tortuous passages is almost entirely in the laminar flow range, and Brownian movement causes impingement and coalescence of the droplets. Each fiber bed is in the form of a hollow cylinder having flow from outside to inside. Clean gas exits the cylinder top and liquid drains from the bottom. Multiple elements keep pressure drop low and, because collection efficiency is largely independent of pressure drop and inlet loading, the collector performs well under start-up and low operating rate conditions where high velocity devices are less efficient.

Elemental phosphorus from the electrothermal process is a distilled product of high purity and yields phosphoric acid pure enough for most industrial uses without any further treatment. The main impurity is ca 20–100 ppm arsenic present in the phosphorus as the element and in the phosphoric acid as arsenious acid. To remove the arsenic, the phosphoric acid destined for food, pharmaceutical, and some industrial-grade applications is treated with excess hydrogen sulfide, filtered, and blown with air to strip out excess H₂S. This treatment generally reduces the arsenic content of the phosphoric acid to less than 0.5 ppm. The small amount of filter cake is disposed of in approved chemical landfills.

The high heat of phosphorus combustion (3053 kJ mol (730 kcal mol) of P₄) is used in the evaporation of water from dilute phosphate solutions that require concentration before subsequent processing. A Progil process, for example, burns phosphorus in a wetted-wall tower in order to concentrate dilute sodium phosphate filtrate from wet-process acid purification (5). The concentration is effected both by evaporation of water and absorption of P₂O₅. Until 1973, the low cost of fuel and the considerable corrosion problems discouraged serious work on the utilization of waste heat for steam (qv) or electric power production (see POWER GENERATION). Rapidly escalating fuel costs, however, have changed this situation. Water-cooled stainless steel heat exchangers, located between the combustion chamber and the hydrator, can be operated below the dew point to recover heat and collect highly concentrated acid (6). Operating above a P₂O₅–H₂O mole ratio of 1.0, the condensed acid is noncorrosive. The heat exchangers may also be operated above the dew point (7).

Wet Process. Over 90% of the phosphoric acid produced, both in the United States and worldwide, is wet-process phosphoric acid used almost exclusively for agricultural application as both fertilizers and animal feed supplements. Although constituting a small proportion of the total wet-acid production, a significant amount of phosphoric acid for food and technical applications is made by purification of wet-process acid.

Wet-process acid is manufactured by the digestion of phosphate rock (calcium phosphate) with sulfuric acid. Depending on availability, other acids such as hydrochloric may be used, but the sulfuric-based processes are by far the most prevalent. Phosphoric acid is separated from the resultant calcium sulfate slurry by filtration. To generate a filterable slurry and to enhance the P₂O₅ content of the acid, much of the acid filtrate is recycled to the reactor. Two main categories of the wet process exist, depending on whether the calcium sulfate is precipitated as the dihydrate or the hemihydrate. Operation at 70–80°C and ~30% P₂O₅ in the liquid phase results in the precipitation of CaSO₄·2H₂O in a filterable form; 80–90°C and ~40% P₂O₅ provide a filterable CaSO₄·0.5H₂O. Operation outside these conditions generally results in poor filtration rates. A typical analysis of wet-process acid is given in Table 4. For more detailed discussion of the wet-process acid, see FERTILIZERS.

Purification. Process development for the purification of wet-process acid has taken place primarily outside North America where the cost differential

between the sulfur used in manufacture of wet-process acid and the electricity needed for thermal acid has been large. Decline of the market for technical-grade phosphates in detergents, along with the escalating cost of electric power for elemental phosphorus production, has resulted in the closing of less efficient elemental phosphorus facilities and the introduction of wet-acid purification into the United States.

Chemical precipitation and solvent extraction are the main methods of purifying wet-process acid, although other techniques such as crystallization (8) and ion exchange (qv) have also been used. In the production of sodium phosphates, almost all wet-process acid impurities can be induced to precipitate as the acid is neutralized with sodium carbonate or sodium hydroxide. The main exception, sulfate, can be precipitated as calcium or barium sulfate. Most fluorine and silica can be removed with the sulfate filter cake as sodium fluorosilicate, Na_2SiF_6 , by the addition of sodium ion and control of the Si/F ratio in the process.

In the double-neutralization process, Na_2SiF_6 is precipitated and removed by filtration at a pH of 3–4 (9). Upon raising the pH to 7–9, insoluble phosphates of Fe, Al, Ca, and Mg form and separate. Iron can be precipitated as hydrous ferric oxide, reducing the phosphate loss at the second filter cake. Both the fluorosilicate and metal phosphate filter residues tend to be voluminous cakes that shrink when dewatered; recovery of soluble phosphates trapped within the cakes is difficult.

The double-neutralization process was used in the production of high volume detergent-builder phosphates, such as sodium tripolyphosphate [7758-29-4] (STP), $\text{Na}_5\text{P}_3\text{O}_{10}$, and tetrasodium pyrophosphate [7722-88-5] (TSP), $\text{Na}_4\text{P}_2\text{O}_7$, because the phosphate precursor solutions occur at pH 7–9. More acidic salts, however, require back-neutralization with pure phosphoric acid, whereas more basic salts may require an additional filtration to eliminate discoloration caused by remaining traces of ferric oxide that precipitate at higher pH. The double-neutralization process is no longer practiced in the United States. More flexibility, better purity, and good economics are provided by modern solvent extraction–purification technology.

Solvent extraction–purification of wet-process phosphoric acid is based on preferential extraction of H_3PO_4 by an organic solvent vs the cationic impurities present in the acid. Because selectivity of acid over anionic impurities is usually not sufficient, precipitation or evaporation steps are included in the purification process for removal. Crude wet-process acid is typically concentrated and clarified prior to extraction to remove post-precipitated sludge and improve partition of the acid into the solvent. Concentration also partially eliminates fluoride by evaporation of HF and/or SiF_4 . Chemical precipitation of sulfate (as Ba or Ca salts), fluorosilicates (as Na salt), and arsenic (as sulfides) may also be used as a prepurification step preceding solvent extraction.

Modern commercial wet-acid purification processes (see Fig. 4) are based on solvents such as C_4 to C_8 alcohols, ethers, ketones, amines, and phosphate esters (10–12). Organic-phase extraction of phosphoric acid is accomplished in one or more extraction columns or, less frequently, in a series of countercurrent mixer–settlers. Generally, 60–75% of the feed acid P_2O_5 content is extracted into the organic phase as H_3PO_4 . The residual phosphoric acid phase (raffinate), containing 25–40% of the original P_2O_5 value, is typically used for fertilizer manufacture such as triple superphosphate. For this reason, wet-acid purification units are almost always located within or next to fertilizer complexes.

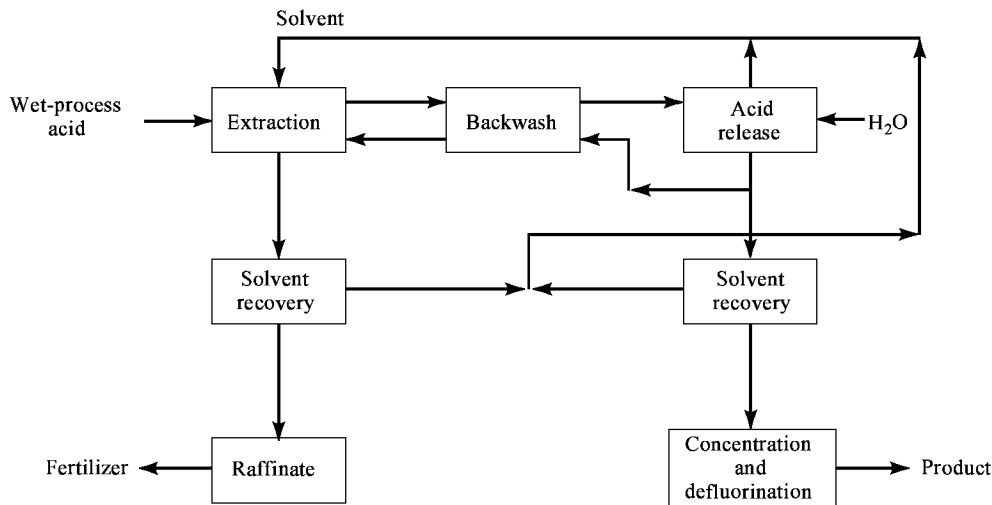


Fig. 4. Schematic diagram of the solvent extraction purification of wet-process phosphoric acid.

The P_2O_5 value in the raffinate is largely in the form of phosphate salts of the metallic impurities which cannot be extracted directly. Higher levels of extraction may result in sludge precipitation in the settlers and the raffinate stream. Higher yields of purified acid may be achieved, however, by the addition of the stronger sulfuric acid for neutralizing the metallic impurities and freeing the residual P_2O_5 value as extractable phosphoric acid. Sulfuric acid may be added either during the extraction step (13), or the raffinate treated in a separate solvent extraction operation. These approaches may pose additional capital and/or energy requirements for phosphoric acid recovery without high levels of sulfate.

The purified acid is recovered from the loaded organic stream by contacting with water in another countercurrent extraction step. In place of water, an aqueous alkali can be used to recover a purified phosphate salt solution. A small portion of the purified acid is typically used in a backwashing operation to contact the loaded organic phase and to improve the purity of the extract phase prior to recovery of the purified acid. Depending on the miscibility of the solvent with the acid, the purified acid and the raffinate may be stripped of residual solvent which is recycled to the extraction loop. The purified acid can be treated for removal of residual organic impurities, stripped of fluoride to low (10 ppm) levels, and concentrated to the desired P_2O_5 level. Many variations of this basic scheme have been developed to improve the extraction of phosphate and rejection of impurities to the raffinate stream, and numerous patents have been granted on solvent extraction processes.

CONDENSED PHOSPHORIC ACIDS

At equilibrium, the specific composition of a concentrated phosphoric acid is a function of its P_2O_5 content. Phosphoric acid solutions up to a concentration equivalent of about 94% H_3PO_4 (68% P_2O_5) contain H_3PO_4 as the only phosphoric acid species present. At higher concentrations, the orthophosphoric acid undergoes condensation (polymerization by dehydration) to yield a mixture of phosphoric acid species (Table 5), often referred to generically as polyphosphoric or superphosphoric acid, $\text{H}_2\text{O} \cdot \text{P}_2\text{O}_5 \sim 3$, or ultraphosphoric acid, $\text{H}_2\text{O} \cdot \text{P}_2\text{O}_5 \sim 1$. At the theoretical P_2O_5 concentration for orthophosphoric acid of 72.4%, the solution is actually a mixture containing 13% pyrophosphoric acid and about 1% free water. Because the pyrophosphoric acid present is the result of an equilibrium state dependent on the P_2O_5 content of the solution, pure orthophosphoric acid can be obtained because of a shift in equilibrium back to H_3PO_4 upon crystallization.

Table 5. Equilibrium Composition of the Strong Phosphoric Acids^a

P ₂ O ₅ , wt % O	P ₂ O ₅ H ₂ O	Percentage composition in terms of the constituent polyphosphoric acids ^b , <i>n</i>														High poly ^c	H ₃ P ₃ O ₉ ^d	H ₄ P ₄ O ₁₂ ^e
		1	2	3	4	5	6	7	8	9	10	11	12	13	14			
67.4	0.263	100.0																
68.7	0.279	99.7	0.33															
70.4	0.302	96.2	3.85															
71.7	0.321	91.0	8.86	f														
73.5	0.352	77.1	22.1	0.79														
73.9	0.360	73.6	25.1	1.34														
75.7	0.394	53.9	40.7	4.86	0.46													
77.5	0.438	33.5	50.6	11.5	2.68	0.74	f											
79.1	0.481	22.1	46.3	20.3	7.82	2.26	1.0	0.3										
							2	4										
80.5	0.523	13.8	38.2	21.0	13.0	6.86	3.3	1.6	1.0	0.2								
							8	7	3	2								
81.0	0.542	12.2	34.0	22.7	14.6	8.42	4.3	2.2	1.4	0.5	f							
							6	7	1	6								
81.2	0.549	10.9	32.9	22.3	15.0	9.36	5.4	2.8	1.7	0.9	0.3	0.0						
							1	5	5	7	6	5						
82.4	0.594	7.32	23.0	19.3	15.9	12.3	8.2	5.7	3.8	2.5	1.3	0.9	0.1			f		
							1	3	9	2	6	1	4					
84.0	0.667	3.92	11.8	12.7	12.0	10.5	8.9	7.9	6.6	5.6	4.5	3.7	3.0	2.4	1.6	6.63		
							7	9	2	3	4	2	3	6	8			
85.0	0.717	2.28	6.36	7.32	8.01	8.17	7.6	7.2	6.9	6.4	5.8	5.2	4.6	3.9	3.8	16.9		
							7	2	3	2	9	7	9	9	3			
85.3	0.736	1.87	4.73	6.33	6.58	6.66	6.7	6.3	6.1	5.8	5.4	5.0	4.9	4.6	4.3	25.6		
							1	6	1	8	6	7	0	4	8			
86.1	0.787	1.46	2.81	3.74	4.43	4.52	4.7	4.7	4.9	4.6	4.5	4.6	4.6	4.3	4.1	43.5	0.17	
							7	9	3	7	4	7	3	8	7			
87.1	0.860	0.83	1.81	2.17	2.53	3.09	3.3	3.4	3.3	3.5	3.4	3.4	3.5	3.2	3.2	61.1	f	
							9	6	3	5	7	5	2	6	4			
87.0	0.920	0.50	0.82	1.56	1.76	1.72	2.0	2.1	2.2	2.0	2.2	2.0	2.2	1.9	2.3	76.4	0.42	0.11
							3	3	6	7	6	6	0	9	0			
89.4	1.066	1.88	1.52	0.77	0.61	0.62	0.6	0.5	0.7	0.8	1.0	0.9	1.1	1.2	1.3	86.8	1.17	0.41
							8	4	1	6	3	8	6	3	7			

^a Ref. 14.

^b *n* = 1, H₃PO₄; 2, H₄P₂O₇; 3, triphosphoric acid [10380-08-2], H₅P₃O₁₀; 4, tetraphosphate; etc.

^c High poly material is retained by resin and includes the phosphoric acid of *n* = 15.

^d Trimetaphosphoric acid [13566-25-1].

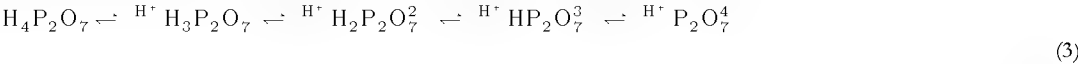
^e Tetrametaphosphoric acid [13598-74-8].

^f Trace.

n=15.

The only clearly defined crystalline compositions are three forms of phosphoric acid and hemihydrate, pyrophosphoric acid, and crystalline P₄O₁₀. The phosphoric acids obtained in highly concentrated solutions or by mixing phosphoric acid with phosphorus pentoxide are members of a continuous series of amorphous (excluding H₄P₂O₇) condensed phosphoric acid mixtures. Mixtures having more than 86% P₂O₅ contain some cyclic metaphosphoric acids [10343-62-1].

Phosphoric acids have one strongly acidic hydrogen for each phosphorus atom of the chain (p*K* ~1–2). At each end of the straight chains, a weakly acidic hydrogen is also present. For long chains, p*K* is ~7.2 for the first of these; 8.2 for the second. The titration curves show inflection points near pH 4.5, 7.5, and 10. Dissociation of pyrophosphoric acid has *K*₁ = 10^{–1}, *K*₂ = 1.5 × 10^{–2}, *K*₃ = 2.7 × 10^{–7}, and *K*₄ = 2.4 × 10^{–10}:



The condensed phosphoric acids are hygroscopic and exhibit a variety of physical forms at room temperature. The material appears oily at 72–82 wt % P₂O₅; viscous and gummy at 82–90 wt % P₂O₅; and is a mixture of glassy and crystalline material at 90 wt %.

Upon boiling dilute phosphoric acid or a collection of P₄O₁₀ in a limited amount of water, an azeotropic mixture is obtained. Its composition varies from 91.1–92.4% P₂O₅ as the pressure of the system increases from 6.6–13.3 kPa to 101.3 kPa (50–100 to 760 mm Hg). The boiling point and composition of vapor over the boiling acid are given in Figure 5 (15). The boiling point increases slowly up to compositions containing 60–70% P₂O₅, and rises rapidly thereafter. The composition of the vapor above boiling phosphoric acid becomes progressively richer in P₂O₅ up to the azeotrope, when both liquid and vapor have an identical composition.

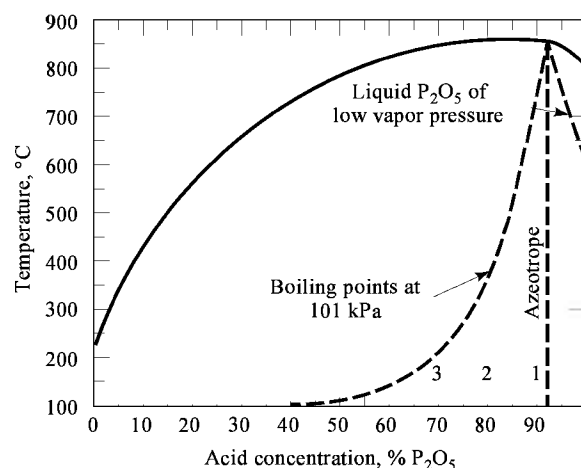


Fig. 5. Vapor-liquid phase diagram of the $\text{H}_2\text{O}-\text{P}_2\text{O}_5$ system at 101 kPa (1 atm), where 3 is ortho, 2 pyro, and 1 meta phosphoric acid. The solid line represents the approximate vapor composition.

Commercial condensed phosphoric acids are mixtures of linear polyphosphoric acids made by the thermal process either directly or as a by-product of heat recovery. Wet-process acid may also be concentrated to $\sim 70\%$ P_2O_5 by evaporation. Linear phosphoric acids are strongly hygroscopic and undergo viscosity changes and hydrolysis to less complex forms when exposed to moist air. Upon dissolution in excess water, hydrolytic degradation to phosphoric acid occurs; the hydrolysis rate is highly temperature-dependent. At 25°C , the half-life for the formation of phosphoric acid from the condensed forms is several days, whereas at 100°C the half-life is a matter of minutes.

Pyrophosphoric (diphosphoric) acid, $\text{H}_4\text{P}_2\text{O}_7$, is the only condensed phosphoric acid definitely obtainable in crystalline form. It has a theoretical P_2O_5 content of 79.8% . However, liquid polyphosphoric acid of such content shows by analysis only 42.5% $\text{H}_4\text{P}_2\text{O}_7$; the remainder is phosphoric acid and various other condensed acids. Pyrophosphoric acid crystallizes in two anhydrous forms. The metastable form, $\text{H}_2\text{P}_2\text{O}_7\text{-I}$, which has an mp of 54.3°C , is usually obtained; the stable form, $\text{H}_2\text{P}_2\text{O}_7\text{-II}$, has an mp of 71.5°C . Pure pyrophosphoric acid solutions are best obtained by ion exchange from the sodium salt. Acid salts are well known.

Traditionally, the term metaphosphoric acid has been used rather freely to describe amorphous mixtures of polyphosphoric acids, especially the azeotropic forms which are also called polymetaphosphoric acids. Commercial forms of polyphosphoric acid are still occasionally referred to in this manner. The term is also used to describe a reagent for the precipitation of protein that contains up to 18% Na_2O in addition to P_2O_5 and H_2O . Technically, the term metaphosphoric acid [10343-62-1] should be reserved for ring-structured acids. Two cyclic acids are reasonably well-defined, tri- and tetrametaphosphoric acids, $\text{H}_3\text{P}_3\text{O}_9$ and $\text{H}_4\text{P}_4\text{O}_{12}$, respectively. Tetrametaphosphoric acid is the main product resulting from the hydrolysis of phosphorus pentoxide and an excess of water in organic medium. Trimetaphosphoric acid can be prepared by ion exchange of its sodium salt, a commercially available material. The cyclic metaphosphoric acids are strong acids having a single, strong inflection in their pH titration curves.

Uses. Owing to extensive use in fertilizers, wet-process phosphoric acid is the largest source of phosphoric acid in the United States, accounting for more than 90% of total acid demand. The remaining phosphoric acid is technical- and food-grade acid supplied by the furnace process or purified wet-process acid. Most of this relatively pure material is marketed in the various forms of phosphate salts. Technical- and food-grade phosphoric acid is used in a variety of applications, including metal treatment, refractories (qv), catalysts, foods, and beverages (see FOOD ADDITIVES).

Cleaners. Phosphoric acid is used in several acidic hard-surface (tile, porcelain, metal) cleaning and sanitizing formulations, as well as an acid cleaner for food processing equipment.

Metal Treatment. After rolling, the oxide scale on sheet steel is removed by acid treatment (pickling) (see METAL SURFACE TREATMENTS). Phosphoric acid, a good pickling agent, leaves the steel coated with a thin film of iron phosphates. This process improves rust resistance but presents a problem if the steel is to be electroplated.

Phosphoric acid-nitric acid baths are used for chemical polishing (bright dipping) of aluminum prior to anodizing. The mixture selectively attacks the metal surface protrusions, resulting in an overall leveling effect. Some copper and brass are also chemically polished with phosphoric acid. Aluminum, steel (including stainless), and other metals are electropolished in relatively high ($50\text{--}80\%$) concentration phosphoric acid baths containing sulfuric and chromic acids as well as other additives.

Iron, zinc, or manganese phosphate coatings are applied to steel-, zinc-, aluminum-, magnesium-, and tin-plated articles to reduce corrosion of the base metal and improve paint adhesion. Such coatings, called phosphate conversion coatings, contain crystalline salts of the metal being treated and the metal ions added to the coating solution. The base metal is attacked by free phosphoric acid. As acid is consumed at the interface, local pH rises, and phosphate salts are precipitated on the base metal, affording the mixed-metal phosphate surface coatings. The zinc phosphate [7779-90-0], $\text{Zn}_3(\text{PO}_4)_2$, coating offers superior paint adhesion, whereas the manganese phosphate provides excellent corrosion resistance (see CORROSION AND CORROSION CONTROL).

Food Additives. Phosphoric acid in dilute solution is nontoxic and has a pleasingly sour taste similar to common food acids such as citric and acetic, but without the fruity flavor of the organic acids. For this reason, it is used widely in cola beverages (see CARBONATED BEVERAGES) as a tart flavoring agent (see FLAVORS AND SPICES). Other food applications include its use as a general protein acidulant, buffering agent in jam and jelly preparation, nutrient and buffer in antibiotic manufacture, acid cleaner for dairy equipment, and purification reagent in sugar refining.

Refractories. Phosphoric acid is used as a bonding agent in various refractory products, particularly alumina, but also magnesia, zirconia, and carbon refractories. Phosphate-bonded refractories typically show improved green strength, load-bearing properties, high temperature stability, and good abrasion resistance.

Condensed Phosphoric Acid. The largest use of polyphosphoric (superphosphoric) acid is as an intermediate in the production of high quality liquid fertilizers. The TVA pioneered the development of electric-furnace superphosphoric acid for this application. However, wet-process superphosphoric acid prepared by evaporation of water from wet-process phosphoric acid has almost completely replaced furnace-grade acid in fertilizer manufacture.

Catalysts. Catalytic applications of phosphoric acid, particularly in the form of condensed thermal polyphosphoric acids, make use of its acidic, nonoxidizing, and dehydrating properties. Condensed acids of $82\text{--}84\%$ P_2O_5 content are employed as catalysts in the petroleum and chemical industries for alkylation, dehydrogenation, polymerization, and isomerization reactions, including the production of adiponitrile from adipic acid, and the manufacture of cumene, ethylbenzene, gasoline, and plasticizer alcohols. The acid is typically supported on diatomaceous earth or other porous materials (see CATALYSTS, SUPPORTED; DIATOMITE). The catalytic activity of polyphosphoric acids depends largely on their hydrogen-ion activity (qv). Dehydration to even higher polyphosphoric acids, with a resulting decrease in strong acidity, begins to occur at around 230°C ; hence, most reactions using polyphosphoric acid catalysts are carried out below this temperature. In addition, steam is often introduced together with the reactants to assure a supply of the more acidic shorter-chain length polyacids via hydrolysis of the more highly condensed acids. Polyphosphoric acid is also used as a dehydrating agent in dye and pigment production. Lesser amounts are used in the production of phosphate esters and agricultural chemicals.

Others. Miscellaneous uses for phosphoric acid are numerous, encompassing wood (qv) and fabric flameproofing, boiler cleaning, opacity control in glass manufacture, textile dyeing, rubber latex coagulation, lithographic engraving, and dental cements (see DENTAL MATERIALS), among others.

Developments in the 1980s and 1990s include phosphoric acid-based formulations for the cleaning of exterior automotive plastic parts and the demand for high purity electronics-grade phosphoric acid.

Phosphates

ORTHOPHOSPHATES

Orthophosphate salts are generally prepared by the partial or total neutralization of orthophosphoric acid. Phase equilibrium diagrams are particularly useful in identifying conditions for the preparation of particular phosphate salts. The solution properties of orthophosphate salts of monovalent cations are distinctly different from those of the polyvalent cations, the latter exhibiting incongruent solubility in most cases. The commercial phosphates include alkali metal, alkaline-earth, heavy metal, mixed metal, and ammonium salts of phosphoric acid. Sodium phosphates are the most important, followed by calcium, ammonium, and potassium salts.

Sodium Phosphates. Elementary chemical considerations might predict three simple sodium phosphates resulting from successive neutralization of the acidic protons of phosphoric acid; ie, monosodium dihydrogen phosphate [7558-80-7] (MSP) NaH_2PO_4 ; disodium monohydrogen phosphate [7558-79-4] (DSP) Na_2HPO_4 ; and trisodium phosphate [7601-54-9] (TSP), Na_3PO_4 . The titration of phosphoric acid with sodium hydroxide shows pH inflections corresponding to two of these three well-known salts (see Fig. 1). The formation of the trisodium salt is too diffuse to be seen. The $\text{Na}_2\text{O}-\text{P}_2\text{O}_5-\text{H}_2\text{O}$ system, actually much more complex than the titration curve indicates, is shown in Figure 6 (16). There are double salts as well as several hydrate forms. Table 6 lists more of these compounds.

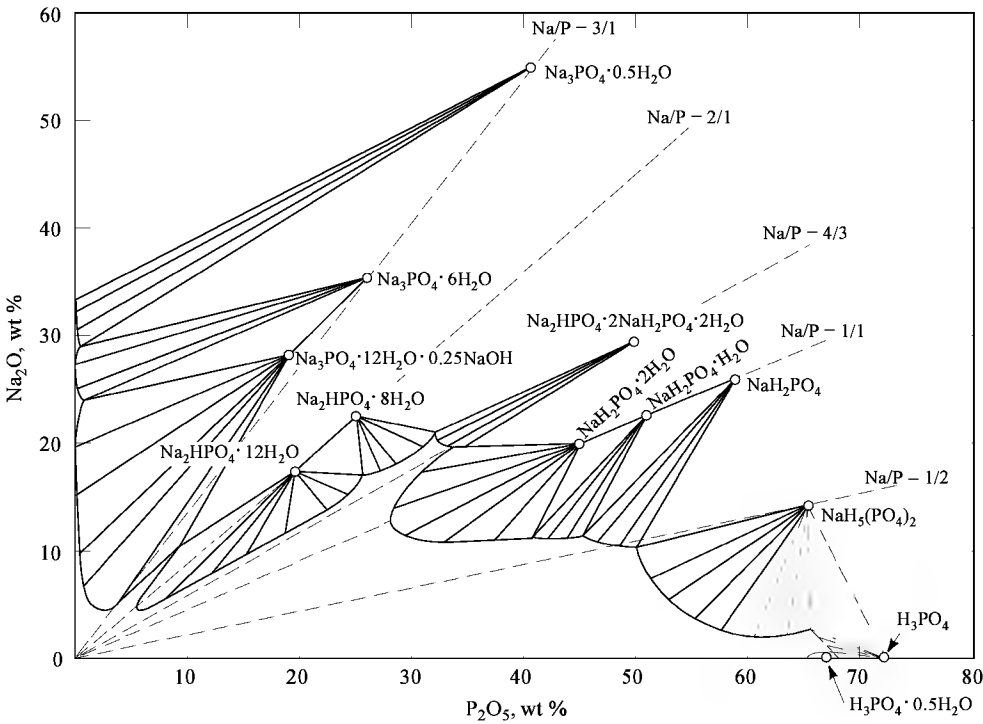


Fig. 6. Phase diagram of the $\text{Na}_2\text{O}-\text{H}_2\text{O}-\text{P}_2\text{O}_5$ (sodium orthophosphate) system at 25°C.

The solubility of a number of sodium orthophosphates is depicted in Figure 7. Some of the sodium phosphates dissolve or precipitate incongruently, affording a different $\text{Na}_2\text{O} / \text{P}_2\text{O}_5$ ratio in the solution phase from that of the solid phase. Sodium phosphates that precipitate are also a function of the temperature. As the temperature increases, the sodium phosphates that may precipitate from solution tend to be anhydrous or lower hydrates than those equilibrium sodium phosphate phases at lower temperature. Whereas most of the phases in Figure 7 represent congruently soluble sodium phosphates, solid phases appear or disappear upon changes in temperature.

Table 6. Sodium Orthophosphates

Compound	CAS Registry Number	Formula
sodium hemiphosphate	[14887-48-0]	$\text{NaH}_2\text{PO}_4 \cdot \text{H}_3\text{PO}_4$
sodium dihydrogen phosphate monohydrate	[10049-21-5]	$\text{NaH}_2\text{PO}_4 \cdot \text{H}_2\text{O}$
sodium dihydrogen phosphate dihydrate	[13472-35-0]	$\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$
sodium dihydrogen phosphate compound with disodium hydrogen phosphate (MSP-DSP)	[39413-44-0]	$\text{NaH}_2\text{PO}_4 \cdot \text{Na}_2\text{HPO}_4$
disodium hydrogen phosphate dihydrate	[10028-24-7]	$\text{Na}_2\text{HPO}_4 \cdot 2\text{H}_2\text{O}$
disodium hydrogen phosphate heptahydrate	[7782-85-6]	$\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$
disodium hydrogen phosphate octahydrate	[67417-37-2]	$\text{Na}_2\text{HPO}_4 \cdot 8\text{H}_2\text{O}$
disodium hydrogen phosphate dodecahydrate	[10039-32-4]	$\text{Na}_2\text{HPO}_4 \cdot 12\text{H}_2\text{O}$
trisodium phosphate hemihydrate	[60593-58-0]	$\text{Na}_3\text{PO}_4 \cdot 0.5\text{H}_2\text{O}$
trisodium phosphate hexahydrate	[15819-50-8]	$\text{Na}_3\text{PO}_4 \cdot 6\text{H}_2\text{O}$
trisodium phosphate octahydrate	[6053-59-1]	$\text{Na}_3\text{PO}_4 \cdot 8\text{H}_2\text{O}$
trisodium phosphate dodecahydrate (TSP crystalline)	[10101-89-0]	$4(\text{Na}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}) \cdot \text{NaOH}$

Both mono- and disodium phosphates are prepared commercially by neutralization of phosphoric acid using sodium carbonate or hydroxide. Crystals of a specific hydrate can then be obtained by evaporation of the resultant solution within the temperature range over which the hydrate is stable. For the preparation of trisodium phosphate, sodium hydroxide must be used to reach the high end pH because CO_2 cannot be stripped readily from the solution above a pH of near 8.

The trisodium phosphate system is the most complex and the commercial product is generally of variable composition and often contains excess sodium hydroxide. It has long been recognized that the usual formula, $\text{Na}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}$, does not accurately represent the constitution of this salt. A better approximation is provided by the formula $4(\text{Na}_3\text{PO}_4 \cdot 12\text{H}_2\text{O}) \cdot \text{NaOH}$, although the exact stoichiometry can vary below this level of sodium. Trisodium phosphate readily forms a variety of double salts with other sodium compounds, generally depicted as $4(\text{Na}_3\text{PO}_4 \cdot 11\text{H}_2\text{O}) \cdot \text{NaCl}$, $5(\text{Na}_3\text{PO}_4 \cdot 11\text{H}_2\text{O}) \cdot \text{NaMnO}_4$, and $4(\text{Na}_3\text{PO}_4 \cdot 11\text{H}_2\text{O}) \cdot \text{NaOCl}$. The double salt of trisodium phosphate and sodium hypochlorite is a source of both alkalinity and active chlorine in disinfectant cleaners and automatic dishwashing formulations. This double salt is generally referred to as chlorinated trisodium phosphate [56802-99-4] (Cl-TSP), although this is a misnomer. It is assumed that Cl-TSP is a mixture of $4(\text{Na}_3\text{PO}_4 \cdot 11\text{H}_2\text{O}) \cdot \text{NaCl}$ and $4(\text{Na}_3\text{PO}_4 \cdot 11\text{H}_2\text{O}) \cdot \text{NaOCl}$ resulting from the chlorination of crystalline trisodium phosphate. However, Cl-TSP contains free sodium chloride and a mixture of phosphate salts. Commercial material approximates the formula $4(\text{Na}_3\text{PO}_4 \cdot 11\text{H}_2\text{O}) \cdot \text{NaOCl}$, and has a typical chlorine content of ca 4%. The pH of a 1% solution is close to 12; water solubility at 25°C is approximately 20 wt %.

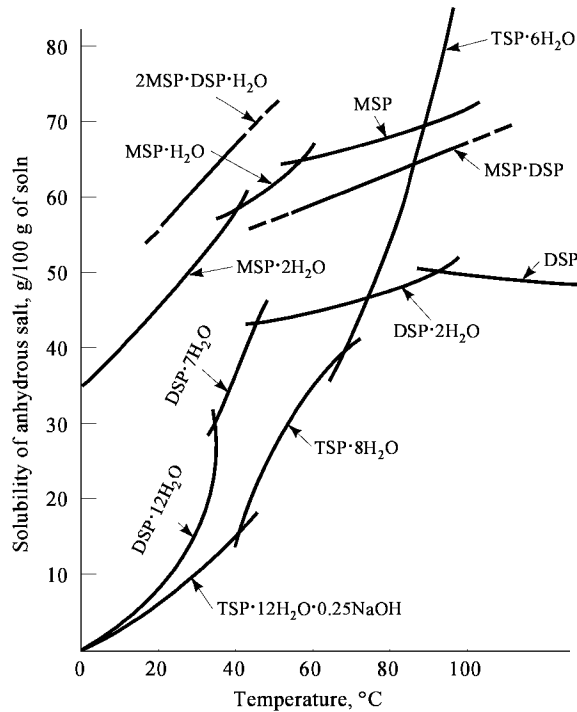


Fig. 7. Solubility of sodium orthophosphates as a function of temperature (17,18). MSP, DSP, and TSP are mono-, di-, and trisodium phosphates, respectively.

Courtesy of Marcel Dekker, Inc.

Uses. The principal use of monosodium phosphate is as a water-soluble solid acid and pH buffer, primarily in acid-type cleaners. The double salt, $\text{NaH}_2\text{PO}_4 \cdot \text{H}_3\text{PO}_4$, referred to as hemisodium orthophosphate or sodium hemiphosphate, is often generated *in situ* from monosodium phosphate and phosphoric acid in these types of formulations. Mixtures of mono- and disodium phosphates are used in textile processing, food manufacture, and other industries to control pH at 4–9. Monosodium phosphate is also used in boiler-water treatment, as a precipitant for polyvalent metal ions, and as an animal-feed supplement.

The single largest use for disodium phosphate is as an emulsifying aid for pasteurized process cheese. Its emulsification mechanism is not completely understood, but the phosphate interacts with the proteins such as casein to allow the proteins to function as emulsifiers and prevent separation of both fat and water in the cheese. A combination with insoluble sodium metaphosphate (IMP) also is used. Typical commercial mixtures contain 30–60% soluble phosphate and 40–70% IMP. Other food-related uses are in ham curing, starch processing, and as an ingredient in instant cereals and evaporated milk (see MILK AND MILK PRODUCTS). Disodium phosphate is also used in the preparation of certain ceramic glazes and enamels, in leather (qv) tanning, textile dyeing, pigment manufacture, water (qv) treatment, and detergents (see ENAMELS, PORCELAIN OR VITREOUS; PIGMENTS; TEXTILES).

Trisodium phosphate is strongly alkaline; many of its applications depend on this property. For example, many heavy-duty cleaning compositions contain trisodium phosphate as a primary alkalinity source. The crystalline dodecahydrate itself is marketed as a cleaning compound and paint remover. Traditionally, trisodium phosphate has been used in water softening to remove polyvalent metal ions by precipitation as insoluble phosphates. Because the hypochlorite complex of trisodium phosphate provides solutions that are strongly alkaline and contain active chlorine, it is used in disinfectant cleaners, scouring powders, and automatic dishwashing formulations.

Potassium Phosphates. The $\text{K}_2\text{O}-\text{P}_2\text{O}_5-\text{H}_2\text{O}$ system parallels the sodium system in many respects. In addition to the three simple phosphate salts obtained by successive replacement of the protons of phosphoric acid by potassium ions, the system contains a number of crystalline hydrates and double salts (Table 7). Monopotassium phosphate (MKP), known only as the anhydrous salt, is the least soluble of the potassium orthophosphates. Monopotassium phosphate has been studied extensively owing to its piezoelectric and ferroelectric properties (see FERROELECTRICS). At ordinary temperatures, KH_2PO_4 is so far above its Curie point as to give piezoelectric effects in which the emf is proportional to the distorting force. There is virtually no hysteresis.

Table 7. Potassium Orthophosphates

Compound	CAS Registry Number	Formula
phosphoric acid, potassium salt (2:1)	[14887-42-4]	$\text{KH}_2\text{PO}_4 \cdot \text{H}_3\text{PO}_4$
potassium dihydrogen phosphate compound with dipotassium hydrogen phosphate monohydrate		$\text{KH}_2\text{PO}_4 \cdot 2\text{K}_2\text{HPO}_4 \cdot \text{H}_2\text{O}$ or $\text{KH}_2\text{PO}_4 \cdot 3\text{K}_2\text{HPO}_4 \cdot \text{H}_2\text{O}$
monopotassium phosphate (MKP)	[7778-77-0]	KH_2PO_4
dipotassium phosphate (DKP)	[7758-11-4]	K_2HPO_4
dipotassium hydrogen phosphate trihydrate	[16788-57-1]	$\text{K}_2\text{HPO}_4 \cdot 3\text{H}_2\text{O}$
dipotassium hydrogen phosphate hexahydrate	[78436-04-1]	$\text{K}_2\text{HPO}_4 \cdot 6\text{H}_2\text{O}$
tripotassium phosphate	[7778-53-2]	K_3PO_4

tripotassium phosphate trihydrate	[22763-03-7]	$\text{K}_3\text{PO}_4 \cdot 3\text{H}_2\text{O}$
tripotassium phosphate heptahydrate	[22763-02-6]	$\text{K}_3\text{PO}_4 \cdot 7\text{H}_2\text{O}$
tripotassium phosphate nonahydrate	[78436-05-2]	$\text{K}_3\text{PO}_4 \cdot 9\text{H}_2\text{O}$

Although the cost of the potassium phosphates is higher than the corresponding sodium salts, the former have applications utilizing their higher solubility and nutrient value. A widespread use for MKP is as a mineral nutrient for fermentation (qv) broths. In a similar vein, special liquid fertilizers contain KH_2PO_4 as a source of both potassium and phosphorus. Monopotassium phosphate is also used in various buffering systems and in paper (qv) processing. The piezoelectric effect of MKP has led to its use in sonar systems and other electronic applications.

Dipotassium phosphate (DKP) and tripotassium phosphate (TKP) are marketed both as solids and in 50% active solution. Most of the commercial output is used in conjunction with borates, nitrites, nitrates, and silicates as the corrosion inhibitor system in ethylene glycol antifreeze formulations (see ANTIFREEZES AND DEICING FLUIDS). The second largest use for DKP is as a buffer in coffee creamers to prevent casein protein coagulation and precipitation by coffee acids. Other uses of DKP include specialty fertilizers, paper-processing aids, and saline cathartics. Tripotassium phosphate is utilized in the polymerization of styrene–butadiene rubber to control the polymerization rate and latex stability. Other applications include rejuvenation of scrap rubber, hydrogen sulfide gas scrubbing, and liquid cleaners.

Ammonium Phosphates. Because ammonium hydroxide is a much weaker base than the common metal hydroxides, and because ammonia may escape as a gas at ambient temperatures, ammonium phosphates are comparatively unstable. For example, both triammonium phosphate [10361-65-6], $(\text{NH}_4)_3\text{PO}_4$, and the double salt [78436-08-5], $(\text{NH}_4)_3\text{PO}_4 \cdot 2(\text{NH}_4)_2\text{HPO}_4$, are unstable at room temperature and evolve ammonia (qv) to form diammonium phosphate [7783-28-0] (DAP), $(\text{NH}_4)_2\text{HPO}_4$. Even the commercial monoammonium phosphate [7722-76-1] (MAP) and DAP exhibit an ammonium vapor pressure both in the solid form and in solution. Although monoammonium phosphate is more stable than diammonium phosphate, which decomposes to ammonia and monoammonium phosphate at around 70°C, MAP also decomposes at high temperature, affording ammonia and polyphosphoric acid. The vapor pressure of a saturated solution of $(\text{NH}_4)_2\text{H}_2\text{PO}_4$ is expressed by the following, when t is between 292 and 363 K.

$$\log_{10} P_{\text{kPa}} = -2240/t + 9.737$$

(4)

When t is between 292 and 328 K, the vapor pressure over a saturated solution of $(\text{NH}_4)_2\text{H}_2\text{PO}_4$ is defined by

$$\log_{10} P_{\text{kPa}} = -2240/t + 9.682$$

(5)

Monoammonium and diammonium phosphates are produced on a large scale as fertilizers. During the 1970s, these materials, produced from economical wet-process phosphoric acid, became the world's leading phosphate fertilizers.

Owing to the thermally unstable nature of ammonium phosphates, other applications are related to flame retarding and fire extinguishing. Monoammonium phosphate is a common fire-extinguishing ingredient in ABC-type dry-power extinguishers. The effectiveness of such extinguishers is closely related to their P_2O_5 content. Ammonium phosphates are used as flame retardants for cellulosic materials, including plywood, papers, and fabrics, to prevent afterglow in matches (qv) and to control forest fires. In flame retarding and fire extinguishing, the phosphoric acid generated during decomposition is thought to catalyze the decomposition of cellulose into a slow-burning char, as well as to smother the flame. Fire retardancy is second only to fertilizers in MAP consumption. Paper, wood, and cloth products not subject to washing are impregnated with an ammonium phosphate solution and dried. Such solutions are often made directly by sparging ammonia into phosphoric acid without crystallizing the solid salts. A high solubility of 140 g/100 g water at 25°C occurs at neutral pH, corresponding to an equimolar mixture of MAP and DAP. Neutral solutions of even higher P_2O_5 content are obtained from anhydrous ammonia and polyphosphoric acids in a short pipeline reactor. The product is cooled immediately to prevent hydrolysis of the polyphosphate.

Evolution of ammonia from a boiling dilute solution of diammonium phosphate gradually reduces the pH. This process is used commercially to control the precipitation of alkali-soluble–acid-insoluble colloidal dyes on wool. Other ammonium orthophosphate salts of interest are ammonium hemiphosphate [28537-48-6], $\text{NH}_4\text{H}_2\text{PO}_4 \cdot \text{H}_3\text{PO}_4$, and its hydrate [28037-74-3], as well as the trihydrate [78436-07-4] of DAP.

Calcium Phosphates. The alkaline-earth phosphates are generally much less soluble than those of the alkali metals. Calcium phosphates include the most abundant natural form of phosphorus, ie, apatites, $\text{Ca}_{10}(\text{PO}_4)_6\text{X}_2$, where $\text{X} = \text{OH}, \text{F}, \text{Cl}$, etc. Apatite ores are the predominant basic raw material for the production of phosphorus and its derivatives. Calcium phosphates are the main component of bones and teeth. After sodium phosphates, the calcium salts are the next largest volume technical- and food-grade phosphates. Many commercial applications of the calcium phosphates depend on their low solubilities.

Several compounds of the $\text{CaO-P}_2\text{O}_5\text{-H}_2\text{O}$ system are given in Table 8. The common names for the mono-, di-, and tricalcium phosphates arise from the traditional double-oxide formulas, $\text{CaO} \cdot 2\text{H}_2\text{O} \cdot \text{P}_2\text{O}_5$, $2\text{CaO} \cdot \text{H}_2\text{O} \cdot \text{P}_2\text{O}_5$, and $3\text{CaO} \cdot \text{P}_2\text{O}_5$, respectively. These terms are routinely used in industry. With the exception of the monocalcium salt, the calcium phosphates are all sparingly soluble.

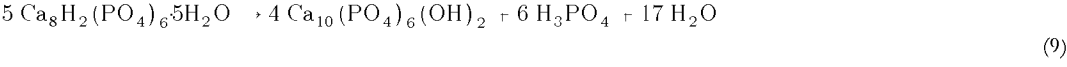
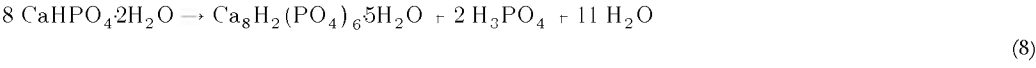
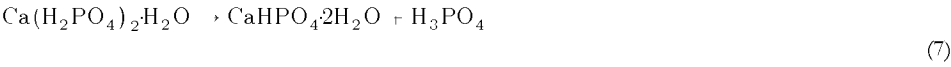
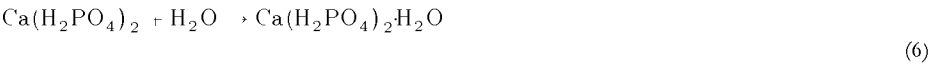
Table 8. Calcium Phosphates in the $\text{CaO-P}_2\text{O}_5\text{-H}_2\text{O}$ System

Compound	CAS Registry Number	Formula	<i>Chemical Abstracts</i> or common name
calcium hydrogen phosphate	[7757-93-9]	CaHPO_4	dicalcium phosphate; monetite [21063-37-6]
calcium hydrogen phosphate hemihydrate	[78436-06-3]	$\text{CaHPO}_4 \cdot 0.5\text{H}_2\text{O}$	metabrushite [78436-06-3]
calcium hydrogen phosphate dihydrate	[7789-77-7]	$\text{CaHPO}_4 \cdot 2\text{H}_2\text{O}$	brushite [14567-92-1]
α-tricalcium phosphate	[7758-87-4]	$\alpha\text{-Ca}_3(\text{PO}_4)_2$	
β-tricalcium phosphate	[7758-87-4]	$\beta\text{-Ca}_3(\text{PO}_4)_2$	
octacalcium phosphate	[14096-86-7]	$\text{Ca}_8\text{H}_2(\text{PO}_4)_6 \cdot 5\text{H}_2\text{O}$	
hydroxyapatite	[1306-06-5]	$\text{Ca}_{10}(\text{PO}_4)_6(\text{OH})_2$	
fluorapatite	[1306-05-4]	$\text{Ca}_{10}(\text{PO}_4)_6\text{F}_2$	
phosphoric acid, calcium salt (2:1)	[7758-23-8]	$\text{Ca}(\text{H}_2\text{PO}_4)_2$	monocalcium phosphate (MCP)
phosphoric acid, calcium salt hydrate (2:1:1)	[10031-30-8]	$\text{Ca}(\text{H}_2\text{PO}_4)_2 \cdot \text{H}_2\text{O}$	monocalcium phosphate monohydrate
phosphoric acid, calcium salt hydrate (2:1:2)	[5221-07-5]	$\text{Ca}(\text{H}_2\text{PO}_4)_2 \cdot 2\text{H}_2\text{O}$	monocalcium phosphate dihydrate

Many orthophosphate salts, in particular those of polyvalent cations, exhibit incongruent solubility where disproportionation occurs in solution to yield a more basic orthophosphate salt and phosphoric acid. This hydrolytic disproportionation of orthophosphates should not be confused with the

hydrolysis of the P–O–P linkage. Incongruent solubility behavior is readily detected by inspection of the aqueous phase diagram for the appropriate system. Hydrolytic disproportionation is probably one of the mechanisms related to the formation of bone and naturally occurring apatites.

Both monocalcium phosphate and dicalcium phosphate dissolve incongruently in water, disproportionating to more basic calcium phosphate and phosphoric acid. The extent of these reactions varies with the temperature and the amount of water. If water is added gradually to anhydrous monocalcium phosphate, equilibrium conditions first correspond to a mixture of the anhydrous salt and its monohydrate. After conversion to the monohydrate, further reaction affords dicalcium phosphate plus free phosphoric acid. Dicalcium phosphate decomposes in aqueous solution to the more basic hydroxyapatite and phosphoric acid via intermediate octacalcium phosphate. The complicated stepwise conversion of the acidic mono- and dicalcium phosphates to hydroxyapatite is summarized in equations 6–9. The kinetics are quite complex.



Because monocalcium phosphate is incongruently soluble, it is typically contaminated with various amounts (6–10%) of dicalcium phosphate and free phosphoric acid resulting from in-process disproportionation of the monocalcium salt. Free phosphoric acid may render the product hygroscopic, and absorbed water plus acid catalyzes further decomposition to additional free acid and dicalcium phosphate. For this reason, industrial monocalcium phosphate may contain some dicalcium phosphate resulting from excess lime addition and then aged to ensure the removal of residual free phosphoric acid.

For many years, the hygroscopic nature of anhydrous monocalcium phosphate limited its commercial applications. However, the addition of small amounts of K⁺, Na⁺, and Al³⁺ ions to the crystallization mother liquor followed by heating to >200°C results in a mixed metal–polyphosphate coating on the calcium phosphate (19). This glassy coating protects the calcium phosphate from moisture and greatly improves its handling properties and performance in several applications, most notably as a leavening agent.

Crystalline CaHPO₄·2H₂O loses both water molecules in a single step at moderately elevated temperature or upon storage to yield the anhydrous salt. The presence of free moisture accelerates this dehydration, which results in anhydrous dicalcium phosphate, often as a hard mass. Addition of a few percent of tetrasodium pyrophosphate or trimagnesium phosphate, Mg₃(PO₄)₂, stabilizes the dihydrate. The mechanism, however, is not well understood. Nonetheless, these materials are used widely to stabilize CaHPO₄·2H₂O, particularly for toothpaste applications.

Tricalcium phosphate, Ca₃(PO₄)₂, is formed under high temperatures and is unstable toward reaction with moisture below 100°C. The high temperature mineral whitlockite [64418-26-4], although often described as β-tricalcium phosphate, is not pure. Whitlockite contains small amounts of iron and magnesium. Commercial tricalcium phosphate prepared by the reaction of phosphoric acid and a hydrated lime slurry consists of amorphous or poorly crystalline basic calcium phosphates close to the hydroxyapatite composition and has a Ca : P ratio of approximately 3:2. Because this mole ratio can vary widely (1.3–2.0), free lime, calcium hydroxide, and dicalcium phosphate may be present in variable proportion. The highly insoluble basic calcium phosphates precipitate as fine particles, mostly less than a few micrometers in diameter. The surface area of precipitated hydroxyapatite is approximately 100 m²/g.

Hydroxyapatite, Ca₁₀(PO₄)₆(OH)₂, may be regarded as the parent member of a whole series of structurally related calcium phosphates that can be represented by the formula M₁₀(ZO₄)₆X₂, where M is a metal or H₃O⁺; Z is P, As, Si, Ga, S, or Cr; and X is OH, F, Cl, Br, 1–2 CO₃, etc. The apatite compounds all exhibit the same type of hexagonal crystal structure. Included are a series of naturally occurring minerals, synthetic salts, and precipitated hydroxyapatites. Highly substituted apatites such as Francolite, Ca₁₀(PO₄)₆(CO₃)_x(F,OH)_{2–x}, are the principal component of phosphate rock used for the production of both wet-process and furnace-process phosphoric acid.

Uses. Commercial monocalcium phosphate is available as both the anhydrous and the monohydrate salts. Most uses are based on acidic properties. Monocalcium phosphate is used to control acidity in powdered drink mixes, as an ingredient in effervescent tablets, as a plastics stabilizer, and in ceramics. Its single largest application is as a leavening agent in bread, cake mixes, and self-rising flour.

The leavening process involves the introduction and expansion of a gas at a controlled rate in a batter or dough system during cooking to afford a light, open texture to the product. Typically, either yeast (see YEASTS) or chemical leavening is employed. Chemical leavening involves the reaction of a dry acid with sodium bicarbonate to release carbon dioxide during batter or dough preparation and cooking. Various chemical leavening acids are available, many of which are phosphate salts, to provide different rates of CO₂ evolution. The phosphate also buffers the pH of the system and interacts with flour proteins to control elasticity and viscosity of the dough (see BAKERY PROCESSES AND LEAVENING AGENTS).

The main use for dicalcium phosphate on a tonnage basis is as an animal feed supplement (see FEEDS AND FEED ADDITIVES), for which it is produced from defluorinated wet-process phosphoric acid. Food-grade dicalcium phosphate is used as a dental polishing agent in toothpastes. This application accounts for the majority of food- and technical-grade production of dicalcium phosphate. The dihydrate is most generally used for this purpose, although anhydrous CaHPO₄ is used in special formulations where more abrasive properties are desired. Dicalcium phosphate also is used as a leavening agent, animal-food mineral supplement, plastics stabilizer, and in the manufacture of glass, medicines, and phosphors.

Commercial tricalcium phosphate is an effective flow conditioner for food products such as sugar and salt. The product is also used as a whitening agent in the manufacture of ceramics, as a mordant in dyeing, and as a polishing agent. Considerable research on apatites has been sparked in the 1980s and 1990s by the desire for biocompatible bone and tooth enamel replacements (see PROSTHETIC AND BIOMEDICAL DEVICES).

Other Orthophosphates. In many instances, magnesium orthophosphates exhibit different properties than the analogous calcium phosphates. Monomagnesium phosphate [13092-66-5], Mg(H₂PO₄)₂, and the di- [15609-80-0] and tetrahydrates [15609-87-7] are somewhat more soluble than the monocalcium phosphates. Unlike dicalcium phosphate dihydrate, dimagnesium phosphate trihydrate [7782-75-4], MgHPO₄·3H₂O, is not incongruently soluble. In the thermal dehydration of dimagnesium phosphate trihydrate, both the hydrate and water and the water of constitution are removed at nearly the same temperature (~140–160°C), generating magnesium pyrophosphate and making the preparation of anhydrous (amorphous) MgHPO₄ in pure form difficult. The more basic magnesium phosphate phases have not been as deeply investigated as for the calcium system. Ammonium magnesium phosphate [7785-21-9], NH₄MgPO₄, is used for gravimetric phosphate analysis because of its insolubility in water.

Aluminum acid phosphates readily form complex polymers when heated above 400°C. Solutions are used as binders in cements and in high temperature bonding of refractories. Aluminum and phosphate ions undergo considerable association in solution. Alumina–phosphoric acid solutions having an Al₂O₃ : P₂O₅ mole ratio of 1.0–1.5 produce highly viscous fluids that, when dried, yield amorphous solids which can be redispersed in water to give solutions stable under acidic conditions. Monoaluminum phosphate [13530-30-2], Al(H₂PO₄)₃, in phosphoric acid solution is employed to surface-treat the steel plates in electrical transformers. Aluminum phosphate [7784-30-7], AlPO₄, is a highly insoluble, hard, and unreactive material with a high melting point (>1800°C) which is used as a refractory material (see REFRACTORIES).

A large number of crystalline phosphates contain two or more cations, and many phosphate minerals are mixed metal salts.

Mixed-sodium–aluminum phosphates are utilized in some food applications, eg, $\text{NaAl}_3\text{H}_{14}(\text{PO}_4)_8 \cdot 4\text{H}_2\text{O}$ and $\text{Na}_3\text{Al}_2\text{H}_{15}(\text{PO}_4)_8$. These salts are prepared by crystallization of a concentrated solution containing the proper $\text{Na}_2\text{O} \text{ } \text{Al}_2\text{O}_3 \text{ } \text{P}_2\text{O}_5$ ratio. The acid sodium aluminum phosphates are used as heat-activated leavening acids, generally in conjunction with the faster-reacting monocalcium phosphate for double-acting baking powders. The sodium aluminum phosphates are of particular interest in flavor-sensitive systems because of their neutral taste. The basic sodium aluminum phosphate [16073-43-1], $\text{Na}_{15}\text{Al}_3(\text{PO}_4)_8$, is used as a food emulsification aid, particularly in processed cheeses (Table 9).

Table 9. Other Phosphates of Commercial Interest

Compound	CAS Registry Number	Formula
aluminum dihydrogen tripolyphosphate	[13939-25-8]	$\text{AlH}_2\text{P}_3\text{O}_{10}$
aluminum phosphate dihydrate (variscite)	[13477-75-3]	$\text{AlPO}_4 \cdot 2\text{H}_2\text{O}$
monoaluminum phosphate sesquihydrate	[78436-09-6]	$\text{Al}(\text{H}_2\text{PO}_4)_3 \cdot 1.5\text{H}_2\text{O}$
dialuminum phosphate trihydrate	[78436-10-9]	$\text{Al}_2(\text{HPO}_4)_3 \cdot 3\text{H}_2\text{O}$
poly(aluminum metaphosphate)	[13776-88-0]	$(\text{Al}(\text{PO}_3)_3)_n$
monoiron(III) phosphate	[18718-09-7]	$\text{Fe}(\text{H}_2\text{PO}_4)_3$
trimagnesium phosphate octahydrate	[13446-23-6]	$\text{Mg}_3(\text{PO}_4)_2 \cdot 8\text{H}_2\text{O}$
aluminum hemiphosphate	[66906-44-3]	$\text{AlH}_3(\text{PO}_4)_2 \cdot \text{H}_2\text{O}$
phosphoric acid, aluminum salt hydrate (2:1:3)	[39611-87-5]	$\text{AlH}_3(\text{PO}_4)_2 \cdot 3\text{H}_2\text{O}$
phosphoric acid, aluminum salt (1:1) hydrate (2:7)	[78436-11-0]	$\text{AlPO}_4 \cdot 3.5\text{H}_2\text{O}$
phosphoric acid, aluminum sodium salt (8:2:3)	[110279-59-1]	$\text{Na}_3\text{Al}_2\text{H}_{15}(\text{PO}_4)_8$
phosphoric acid, aluminum sodium salt (3:3:1) tetrahydrate	[10305-76-7]	$\text{NaAl}_3\text{H}_{14}(\text{PO}_4)_8 \cdot 4\text{H}_2\text{O}$
phosphoric acid, iron(III) salt hydrate (2:1:2.5)	[31359-25-8]	$\text{FeH}_3(\text{PO}_4)_2 \cdot 2.5\text{H}_2\text{O}$
triposphoric acid, monosodium salt	[15575-11-8]	$\text{NaH}_4\text{P}_3\text{O}_{10}$
triposphoric acid, disodium salt	[33689-84-8]	$\text{Na}_2\text{H}_3\text{P}_3\text{O}_{10}$
triposphoric acid, trisodium salt	[13772-25-3]	$\text{Na}_3\text{H}_2\text{P}_3\text{O}_{10}$
triposphoric acid, tetrasodium salt	[25616-37-3]	$\text{Na}_4\text{HP}_3\text{O}_{10}$
pentasodium salt	[7758-29-4]	$\text{Na}_5\text{P}_3\text{O}_{10}$
sodium potassium tripolyphosphate	[24315-83-1]	$\text{Na}_5\text{P}_3\text{O}_{10} \cdot \text{K}_5\text{P}_3\text{O}_{10}$
sodium trimetaphosphate	[7785-84-4]	$\text{Na}_3\text{P}_3\text{O}_9$
sodium tetrametaphosphate	[13396-41-3]	$\text{Na}_4\text{P}_4\text{O}_{12}$
sodium hexametaphosphate	[10124-56-8]	$(\text{NaPO}_3)_n^a$
poly(sodium metaphosphate) (insoluble metaphosphate (IMP))	[10361-03-2]	$(\text{NaPO}_3)_n$
zirconium phosphate monohydrate	[13933-56-7]	$\text{Zr}(\text{HPO}_4)_2 \cdot \text{H}_2\text{O}$
zirconium phosphate dihydrate	[13772-31-1]	$\text{Zr}(\text{HPO}_4)_2 \cdot 2\text{H}_2\text{O}$

^a $n = \sim 6\text{--}20$.

Iron phosphates are generally similar to aluminum phosphates. The two varieties of $\text{Fe}(\text{H}_2\text{PO}_4)_3$ are isomorphous with the two forms of $\text{Al}(\text{H}_2\text{PO}_4)_3$. Both varieties of $\text{Fe}(\text{H}_2\text{PO}_4)_3$ are highly hygroscopic and hydrolyze to $\text{FeH}_3(\text{PO}_4)_2 \cdot 2.5\text{H}_2\text{O}$ and an acidic solution. Commercial applications of the iron phosphates are quite limited but include catalysts, mineral supplements, and specialty glass manufacture.

Zinc phosphate, $\text{Zn}_3(\text{PO}_4)_2$, forms the basis of a group of dental cements. Chromium and zinc phosphates are utilized in some metal-treating applications to provide corrosion protection and improved paint adhesion. Cobalt(II) phosphate octahydrate [10294-50-5], $\text{Co}_3(\text{PO}_4)_2 \cdot 8\text{H}_2\text{O}$, is a lavender-colored substance used as a pigment in certain paints and ceramics. Copper phosphates exhibit bioactivity and are used as insecticides and fungicides. Zinc, lead, and silver phosphates are utilized in the production of specialty glasses. The phosphate salts of heavy metals such as Pb, Cr, and Cu, are extremely water insoluble.

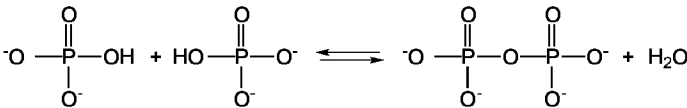
The tertiary metal phosphates are of the general formula MPO_4 , where M is B, Al, Ga, Fe, Mn, etc. The metal–oxygen bonds of these materials have considerable covalent character. The anhydrous salts are continuous three-dimensional networks analogous to the various polymorphic forms of silica. Of limited commercial interest are the aluminum, boron, and iron phosphates. Boron phosphate [13308-51-5], BPO_4 , is produced by heating the reaction product of boric acid and phosphoric acid or by adding H_3BO_3 to H_3PO_4 at room temperature, followed by crystallization from a solution containing >48% P_2O_5 . Boron phosphate has limited use as a catalyst support, in ceramics, and in refractories.

Many phosphates exhibit two- or three-dimensional structures (20). The titanium and zirconium phosphates, $\text{M}(\text{HPO}_4)_2 \cdot n\text{H}_2\text{O}$, form inert, high temperature-stable ion-exchange agents possessing a layered structure. Both $\alpha\text{-Zr}(\text{HPO}_4)_2 \cdot \text{H}_2\text{O}$ and $\gamma\text{-Zr}(\text{HPO}_4)_2\text{O}$ are of particular interest. Proposed uses include high temperature processing of nuclear waste and kidney dialysis. Polar organics may also be intercalated between the layers, and reactions such as phosphate group exchange can be catalyzed. The Nasicon family of compounds, $\text{Na}_{1-x}\text{Zr}_2\text{Si}_x\text{P}_{3-x}\text{O}_{12}$, are three-dimensional sodium ion conductors that also have controllable, near-zero thermal expansion properties.

CONDENSED PHOSPHATES

Condensed phosphates are derived by dehydration of acid orthophosphates. The resulting polymeric structures are based on a backbone of P–O–P linkages where PO_4 tetrahedra are joined by shared oxygen atoms. The range of materials within this classification is extremely broad, extending from the simple diphosphate, also known as pyrophosphate, to indefinitely long-chain polyphosphates and ultraphosphates (see Table 1). Both well-defined crystalline and amorphous materials occur among the condensed phosphates.

Pyrophosphates. The simplest linear condensed phosphates are pyrophosphates, which can be considered as the dehydration product of two orthophosphate groups. A water molecule is eliminated to form a P–O–P linkage in a reversible reaction.



Many pyrophosphates can be prepared by thermal treatment of the acid orthophosphates.



Compositions having an oxide ratio, $R < 2$, eg, in equation 11, can dehydrate to form more highly condensed phosphates beyond the pyrophosphate. Conditions must therefore be controlled to prevent complete dehydration. Insoluble pyrophosphates are obtained by treatment of a soluble salt of the desired cation using a sodium pyrophosphate solution.

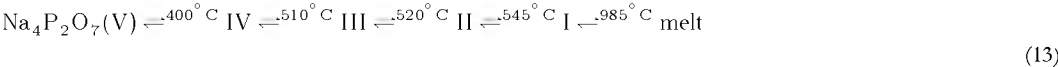
Most crystalline pyrophosphate salts have a nonlinear configuration; the angle of the central P–O–P linkage varies with different cations. Most anhydrous pyrophosphates are thermally stable up to their melting points, although upon heating many undergo polymorphic phase changes in crystalline structure involving an alteration of the P–O–P angle.

A large number of pyrophosphate salts have been prepared (Table 10). In addition to individual metal salts, ammonium pyrophosphates and many mixed-metal pyrophosphates are known. Pyrophosphates of notable commercial importance include sodium, potassium, and calcium salts.

Table 10. Pyrophosphates

Compound	CAS Registry Number	Formula
aluminum pyrophosphate	[14696-66-3]	Al ₄ (P ₂ O ₇) ₃
calcium dihydrogen pyrophosphate (calcium acid pyrophosphate)	[14866-19-4]	CaH ₂ P ₂ O ₇
calcium pyrophosphate	[7790-76-3]	Ca ₂ P ₂ O ₇
potassium trihydrogen pyrophosphate	[16270-75-0]	KH ₃ P ₂ O ₇
dipotassium dihydrogen pyrophosphate (potassium acid pyrophosphate)	[14691-84-0]	K ₂ H ₂ P ₂ O ₇
tripotassium hydrogen pyrophosphate	[16270-76-1]	K ₃ HP ₂ O ₇
tetrapotassium pyrophosphate	[7320-34-5]	K ₄ P ₂ O ₇
sodium trihydrogen pyrophosphate (monosodium pyrophosphate)	[13847-74-0]	NaH ₃ P ₂ O ₇
disodium dihydrogen pyrophosphate (sodium acid pyrophosphate)	[7758-16-9]	Na ₂ H ₂ P ₂ O ₇
disodium dihydrogen pyrophosphate hexahydrate	[13510-98-0]	Na ₂ H ₂ P ₂ O ₇ ·6H ₂ O
trisodium hydrogen pyrophosphate (trisodium pyrophosphate)	[14691-80-6]	Na ₃ HP ₂ O ₇
trisodium hydrogen pyrophosphate monohydrate	[26573-04-6]	Na ₃ HP ₂ O ₇ ·H ₂ O
trisodium hydrogen pyrophosphate nonahydrate	[16457-94-6]	Na ₃ HP ₂ O ₇ ·9H ₂ O
tetrasodium pyrophosphate (TSPP)	[7722-88-5]	Na ₄ P ₂ O ₇
tetrasodium pyrophosphate decahydrate	[13472-36-1]	Na ₄ P ₂ O ₇ ·10H ₂ O
silicon pyrophosphate	[13827-38-8]	SiP ₂ O ₇
titanium pyrophosphate	[13470-09-2]	TiP ₂ O ₇

Sodium Pyrophosphates. Known pyrophosphate compounds in the Na₂O–H₂O–P₂O₅ system are given in Table 10. Commercially important sodium pyrophosphates include tetrasodium pyrophosphate (TSPP), Na₄P₂O₇, and disodium pyrophosphate, Na₂H₂P₂O₇, commonly referred to as sodium acid pyrophosphate (SAPP). These are prepared industrially by thermal dehydration of disodium and monosodium orthophosphate, respectively. Tetrasodium pyrophosphate exists in five crystalline modifications, only one of which is stable at room temperature.



TSPP is readily crystallized from water as the decahydrate between −0.4° and 79°C, and as the anhydrous salt above 79°C. The solubility of tetrasodium pyrophosphate is illustrated in Figure 8. The pH of a 1% solution is 10.2. TSPP is quite stable in alkaline medium but hydrolyzes rapidly to orthophosphate under acid conditions.

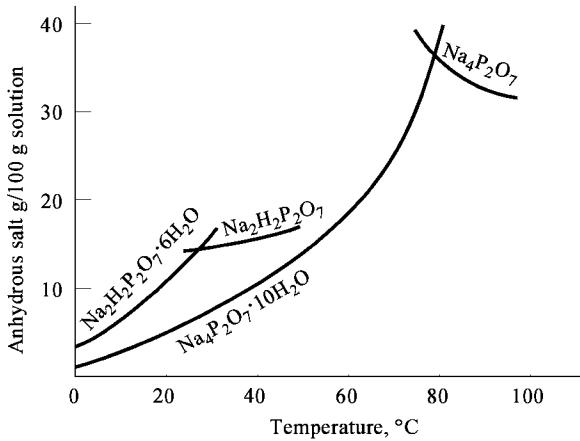


Fig. 8. Solubilities of sodium acid pyrophosphate and tetrasodium pyrophosphate.

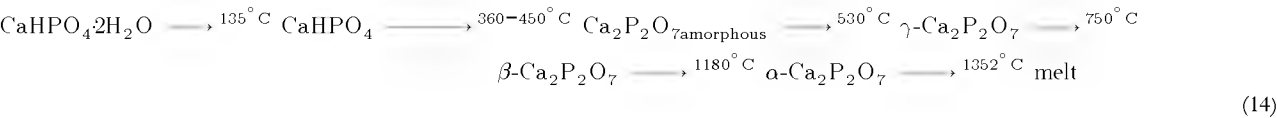
Tetrasodium pyrophosphate is a builder in detergent and cleaning formulations. Food applications include consistency control in buttermilk (thinning), chocolate milk (thickening), and instant puddings. Thinning action results from neutralization of lactic acid that agglomerates butterfat; thickening results from the formation of calcium pyrophosphate gels. Tetrasodium pyrophosphate is widely used as an effective deflocculant, eg, in kaolin clays, drilling muds, dyes, and inks. TSPP stabilizes hydrogen peroxide through chelation of heavy-metal ion impurities that catalyze peroxide decomposition. TSPP is used as an anticalculus agent in toothpastes and mouthwashes (see DENTIFRICES). A primary component of calculus is a calcium scale and TSPP is thought to function by complexing soluble calcium ion and inhibiting crystallization from the tooth plaque (21).

The solubility of Na₂H₂P₂O₇ is also illustrated in Figure 8. The hexahydrate, Na₂H₂P₂O₇·6H₂O, is the solid form in equilibrium with solution up to near 27°C, whereupon the anhydrous form is obtained above this temperature. Commercial Na₂H₂P₂O₇ typically contains small amounts of added

potassium, calcium, and aluminum to reduce the rate of reaction with sodium bicarbonate in leavening, the largest application of SAPP. More reactive forms of SAPP are particularly suited for making doughnuts, whereas the slower reacting grades are used in refrigerated canned dough products such as biscuits. However, SAPP sometimes imparts an astringent aftertaste to certain baked goods. SAPP is also used to eliminate darkening of cut potatoes to prevent the formation of highly colored iron–tannin complexes. The pyrophosphate chelates the iron as a colorless iron pyrophosphate. Other applications are in acid-cleaning formulations and in electroplating.

Potassium Pyrophosphates. Tetrapotassium pyrophosphate (TKPP), $K_4P_2O_7$, is easily prepared by thermal dehydration of K_2HPO_4 . TKPP is highly soluble in water, 187.4 g / 100 g H_2O at 25°C. In a saturated solution, the equilibrium solid is $K_4P_2O_7 \cdot 3H_2O$. Potassium pyrophosphates are typically quite hygroscopic. Tetrapotassium pyrophosphate has been used as a highly soluble detergent builder because it permits easy formulation in liquid detergent systems. Unlike TKPP or SAPP, dipotassium dihydrogen pyrophosphate, $K_2H_2P_2O_7$, is difficult to prepare by thermal dehydration of KH_2PO_4 because the rate of conversion to the fully dehydrated metaphosphate composition, KPO_3 , is fast.

Calcium Pyrophosphates. As is typical of the pyrophosphate salts of multiple-charged or heavy-metal ions, the calcium pyrophosphates are extremely insoluble in water. Calcium pyrophosphate exists in three polymorphic modifications, each of which is metastable at room temperature. These are formed progressively upon thermal dehydration of calcium hydrogen phosphate dihydrate as shown below. Conversion temperatures indicated are those obtained from thermal analyses (22,23). The presence of impurities and actual processing conditions can change these values considerably, as is true of commercial manufacture.



The calcium pyrophosphates are utilized primarily as dental abrasives in fluoride-containing toothpastes. Typically, a mixture of β - and γ - $Ca_2P_2O_7$ achieves a balance of abrasivity and fluoride compatibility.

Tripolyphosphates. The most commercially important tripolyphosphate salt is sodium tripolyphosphate (STP), $Na_5P_3O_{10}$. Three distinct crystalline forms are known: two are anhydrous (STP-I and STP-II); the other is the hexahydrate [15091-98-2], $Na_5P_3O_{10} \cdot 6H_2O$. Sodium tripolyphosphate anhydrous Form I is the high temperature, thermodynamically stable phase; sodium tripolyphosphate anhydrous Form II is the lower temperature form which can be readily converted to STP-I by heating to above $417 \pm 8^\circ C$, the transition temperature. However, the reverse reaction is extremely slow below 417°C. Both anhydrous forms of sodium tripolyphosphate are therefore stable enough to coexist at room temperature.

The structures of STP-I and -II differ primarily in the ionic coordination of cations. In STP-II all sodium ions are octahedrally coordinated by oxygen, whereas in STP-I some sodium ions are surrounded by only four oxygen atoms. In STP-II a distinct sheet-like arrangement occurs. The faster hydration rates are attributed to these properties.

The hexahydrate is formed by the addition of anhydrous STP to water or by the hydrolysis of sodium trimetaphosphate [7785-84-4] (STMP), $(NaPO_3)_3$, in alkaline media. The hexahydrate is stable at room temperature but undergoes rapid hydrolytic degradation to pyro- and orthophosphate upon drying.

Sodium tripolyphosphate is produced by calcination of an intimate mixture of orthophosphate salts containing the correct overall Na / P mole ratio of 1.67. The proportions of the two anhydrous STP phases are controlled by the calcination conditions. Commercial STP typically contain a few percent of tetrasodium pyrophosphate and some trimetaphosphate. A small amount of unconverted orthophosphates and long-chain polyphosphates also may be present.

The two anhydrous forms of sodium tripolyphosphate can be differentiated by x-ray diffraction, or by infrared or Raman spectroscopy. For industrial purposes, the temperature rise (TR) test (24) is based on the faster hydration rate of STP-I in a water–glycerol mixture, where % STP-I = 4 (TR – 6). Low TR (LTR) STP has a TR value of 6–10 and is predominantly Form II STP. High TR (HTR) STP has a TR range of 10–20 and is a mixture of Forms I and II. Ultrahigh TR (UHTR) STP consists almost entirely of Form I STP and has a typical TR value of 25 or higher.

The solubility and hydration behavior of sodium tripolyphosphate are of particular importance in many of its industrial applications. At room temperature, $Na_5P_3O_{10} \cdot 6H_2O$ dissolves to an equivalent $Na_5P_3O_{10}$, of about 13 g / 100 g solution. Both anhydrous forms are more soluble than the hexahydrate but each is unstable with respect to the hexahydrate. STP-I is the more rapidly hydrating form having a higher transient solubility than STP-II. Dissolution of STP-I is accompanied by rapid formation of crystalline hexahydrate, and the higher initial solubility drops almost immediately to that of STP $\cdot 6H_2O$. STP-II, on the other hand, dissolves less readily in water, but supersaturated solutions containing 35% or more STP-II can be formed from which the hexahydrate crystallizes slowly. Thus, for both forms of anhydrous STP, maximum solubility under a given set of conditions is determined by two opposing factors, the rate of the solution of anhydrous salt, and the rate of the crystallization of hexahydrate. The greater ease of STP-I hydration is attributed to the lower degree of coordination for the sodium ions in its crystalline structure. Phosphate impurities in STP affect the solubility and hydration behavior. For example, the presence of a small amount of glassy polyphosphate can stabilize supersaturated solutions of STP-II for several hours. STP may be hydrated during manufacture to ca 0.5–1% moisture to form some portion as the hexahydrate. The presence of seed hexahydrate crystals promotes rapid hydration and little tendency to supersaturate. STP particle size, order of addition, agitation, etc, can also affect properties such as the dissolution rate, clumping of solid, and the persistence of supersaturation.

The hydration rate of sodium tripolyphosphate to its stable hexahydrate, $Na_5P_3O_{10} \cdot 6H_2O$, directly affects detergent processing and product properties. The proportion of STP-I (fast-hydrating form) and STP-II (slow-hydrating form) in commercial sodium tripolyphosphate is controlled by the time–temperature profile during calcination. In most processes, a final product temperature of near 450°C results in a product containing about 30% STP-I, which is desired by detergent manufacturers. Addition of a small amount of water to the sodium tripolyphosphate furnishes hexahydrate seed crystals that minimize any induction period or variation in hydration in the detergent manufacturing process. Water is added during STP manufacture either after cooling and before milling, by atomizing water into special high intensity blenders, or by vapor-phase hydration.

The pH of a 1% solution of pure sodium tripolyphosphate is 9.9 and that of commercial samples may vary between 9.5 and 10.1. The pH of a given sample of solid STP drops slowly with age because of water adsorption and partial reversion to orthophosphate and pyrophosphate. The pH of solutions varies with concentration because the sodium ion is bound in the complex form $NaP_3O_4^{10-}$ at higher concentrations; maximum pH is reached at between 1–2% solution.

Anhydrous sodium tripolyphosphate is slow to hydrate in contact with the atmosphere under normal ambient conditions and generally remains free-flowing. If the relative humidity is below a critical relative humidity, which is different for both anhydrous forms of STP and dependent on temperature, hydration does not take place. For prolonged storage at room temperature, relative humidities above ca 60% in the air result in water absorption. For shorter periods, high levels of humidity can be tolerated. However, even at higher humidities, the amount of water absorbed is small. The heats evolved from vapor hydration of STP-I and -II have been estimated at 343 and 334 kJ / mol (82.0 and 79.9 kcal / mol), respectively (25).

Uses. Sodium tripolyphosphate was introduced in the 1940s as a builder for synthetic detergents. It was once the largest volume commercial product manufactured from technical-grade phosphoric acid but in the 1990s volumes are small compared to the nearly 1×10^6 kg produced in the United States alone in 1970. As a builder in cleaning formulations, sodium tripolyphosphate is used in household laundry products, automatic dishwashing formulations, car washes, and numerous industrial cleaners. STP provides many functions to improve the cleaning efficiency, including the sequestration of hardness ions, buffered alkalinity, deflocculation of dirt particles, and antiredeposition of soil.

Food-grade sodium tripolyphosphate is used for the curing of hams and bacon, where interaction with the meat protein allows for better moisture retention during cooking. Treatment with STP improves the quality of poultry and seafood products. Uses for technical-grade material include clay processing, water softening, textile processing, paper pulping, rubber and paint manufacture, drilling muds, and ore flotation.

Other Tripolyphosphates. Potassium tripolyphosphate [24315-83-1] (KTP), $K_5P_3O_{10}$, has a high aqueous solubility (near 180 g/100 g) and has been used in place of STP for liquid detergents. The potassium salt, however, is more expensive than STP. Sodium potassium tripolyphosphate (SKTP), $Na_3K_2P_3O_{10}$, is prepared by calcination of a feed liquor having the proper $Na_2O/K_2O/P_2O_5$ ratio or by reaction of sodium trimetaphosphate with KOH. For some detergent or food applications, SKTP may provide the optimum compromise between solubility and cost.

Long-Chain Polyphosphates and Metaphosphates. Polyphosphates larger than tetrapolyphosphate are difficult to obtain in pure form. These usually occur as amorphous glassy materials having a distribution of various chain lengths. As the chains become long, however, the properties of the individual chains become so similar that the mixtures behave much like pure compounds and occur as crystalline substances. The composition of long-chain polyphosphates approaches that of metaphosphate, $(MPO_3)_n$, and long-chain polyphosphates may commonly be referred to as metaphosphates, although this term should be reserved for cyclic anions of the exact $(PO_3)_n$ composition. Most polyphosphates are amorphous glasses, but several high molecular weight polyphosphates occur as crystalline substances. Both types are used commercially.

Thermal dehydration of monosodium phosphate gives rise to numerous condensed polyphosphates (Fig. 9). Structures are diverse and can be controlled by manipulating the conditions of dehydration, ie, temperature, water vapor, and tempering. Graham's salt is a water-soluble polyphosphate glass having a composition approximating $(NaPO_3)_{50}$. It is manufactured by heating NaH_2PO_4 to above 620°C and quenching rapidly. The melt solidifies into a glass consisting of about 90% high molecular weight polyphosphates. The remainder is a mixture of various cyclic metaphosphates. One example of a glassy phosphate has a Na_2O/P_2O_5 ratio of about 1.05–1.1, a molecular weight of 1500–2000, and a degree of polymerization of 15–20. Similar commercial materials of average chain lengths of 6–8, 12–14, and up to ~30 are available. The misnomer hexametaphosphate is often applied to the glassy polyphosphates of this type. Glassy polyphosphates are used in water treatment, owing to the ability of the polyphosphate anion to sequester hardness ions (Ca, Mg, Fe) and convert them into soluble complexes. Polyphosphates are effective thresholding agents, because ppm levels inhibit scale formation in water systems. Glassy polyphosphates act as dispersants on finely divided solids in clay processing, oil-well drilling muds, and pigment manufacture.

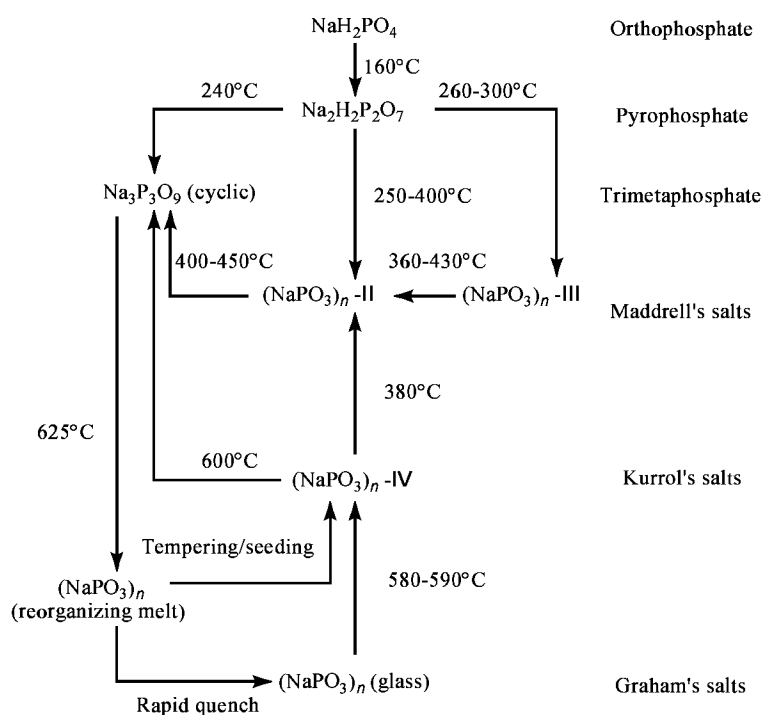
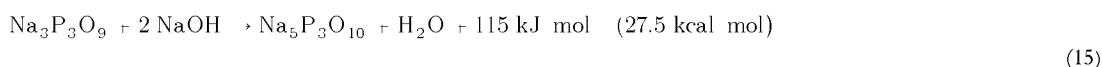


Fig. 9. Sodium polyphosphate conversions.

Several crystalline condensed phosphates may also be formed by the dehydration of monosodium phosphate (MSP). Maddrell's salt exists as Form II (high temperature Maddrell, $NaPO_3$ -II, insoluble metaphosphate-II) and Form III (low temperature Maddrell, $NaPO_3$ -III, insoluble metaphosphate-III). Both forms are highly polymerized and difficult (slow) to dissolve in water. Mixtures of the two forms are marketed as a dental abrasive for toothpaste formulations containing soluble fluoride. Maddrell's salt is also used with disodium phosphate as a cheese emulsifying aid.

The only true metaphosphate (ring structure) of significant commercial interest is sodium trimetaphosphate (STMP), $Na_3P_3O_9$. Because of the strain inherent in the small ring structure, STMP is more reactive toward nucleophiles than chain phosphates. In the presence of NaOH, for example, STMP forms sodium tripolyphosphate.



This reaction is the basis for the Fluff detergent process (26). The heat evolved coupled with the hydration of the STP is sufficient to dry the reaction mass to yield a light density laundry detergent. Requiring a low capital investment, the Fluff process has found use in less-industrialized countries.

STMP reacts with other nucleophiles such as aqueous ammonia to yield amidophosphates, which contains a P–N bond. STMP is used for the modification of the physical properties of starch and proteins by reaction with the amino and hydroxyl groups.

Potassium Kurrol's salt, potassium polymetaphosphate [7790-53-6], $(KPO_3)_n$, is easily obtained by thermal dehydration of KH_2PO_4 . The degree of polymerization is directly related to the K_2O/P_2O_5 ratio, temperature, and length of heating. Kurrol's salt typically has a high molecular weight, perhaps up to several million. Sodium Kurrol's salt is a fibrous, crystalline material and, unlike the potassium analogue, is difficult to crystallize in preference to other sodium poly- or metaphosphates. Both sodium and potassium Kurrol's salt are slowly soluble in water but readily dissolve to form viscous solutions in the presence of other alkali-metal cations. The potassium salt has limited commercial usage in sausage processing in Europe.

Ammonium polyphosphate [13446-46-3], $(NH_4PO_3)_n$, can be produced directly by thermal dehydration of $NH_4H_2PO_4$ if care is taken to maintain a high ammonia pressure over the system; otherwise the predominant reaction is the loss of ammonia. It is most easily obtained, however, by heating a mixture of $NH_4H_2PO_4$ and urea in an atmosphere of NH_3 . Alternative methods include ammoniation of polyphosphoric acid or condensation of ammonium tetrametaphosphate [14693-64-2], $(NH_4PO_3)_4$, a product recovered by stirring P_4O_{10} with concentrated ammonium hydroxide. There are at least five crystalline forms under the generic term ammonium polyphosphate. Form I, NH_4PO_3 -I, is used as a water-insoluble fire retardant in intumescent paints and mastics. Form II, having a higher temperature stability than Form I, is used as a fire retardant in thermoplastics. Ammonium polyphosphate liquid fertilizers are made from wet-process acid.

Properties of Condensed Phosphates.

Hydrolysis. Condensed phosphates all exhibit hydrolytic instability of the P–O–P linkages and, under the appropriate conditions, can all be

cleaved, ultimately affording the monomeric orthophosphate ion (see eq. 10). Aqueous phase diagrams assume metastability for condensed phosphates and ignore hydrolysis to orthophosphates, although limiting the hydrolytic degradation is an important experimental consideration in defining the phase diagram. Like the orthophosphates, condensed phosphates may also exhibit hydrolytic disproportionation, ie, incongruent solubility.

The hydrolysis rates of polyphosphates are mainly affected by temperature, pH, and the location of the phosphate group in a condensed phosphate. P–O–P linkages attached to triply connected phosphate groups (ultraphosphates) undergo extremely rapid hydrolysis in aqueous systems. Pyrophosphate, containing only end (singly connected) phosphate groups, is notably more stable toward hydrolytic degradation than chain phosphates, which also contain middle (doubly connected) phosphate moieties.

Hydrogen ion is a good catalyst for hydrolytic degradation of the P–O–P linkage. Hydrolysis of short-chain phosphates is extremely slow at room temperature and neutral pH. However, raising the temperature and or lowering the pH can increase the hydrolysis rate. In dilute and moderately concentrated solution, the overall kinetics for hydrolysis is first-order at a given pH. An estimation of hydrolysis rates can be obtained from the nomograph of Figure 10. The rupture of a P–O–P linkage generates acid end groups, causing the pH of phosphate solutions to fall as hydrolysis occurs and is therefore autocatalytic. Activation energies decrease from ca 125–165 kJ mol (30–40 kcal mol) at pH 10 to ca 85–125 kJ mol (20–30 kcal mol) at pH 4–7 for sodium pyro- and tripolyphosphates, with values dropping in the presence of catalysts.

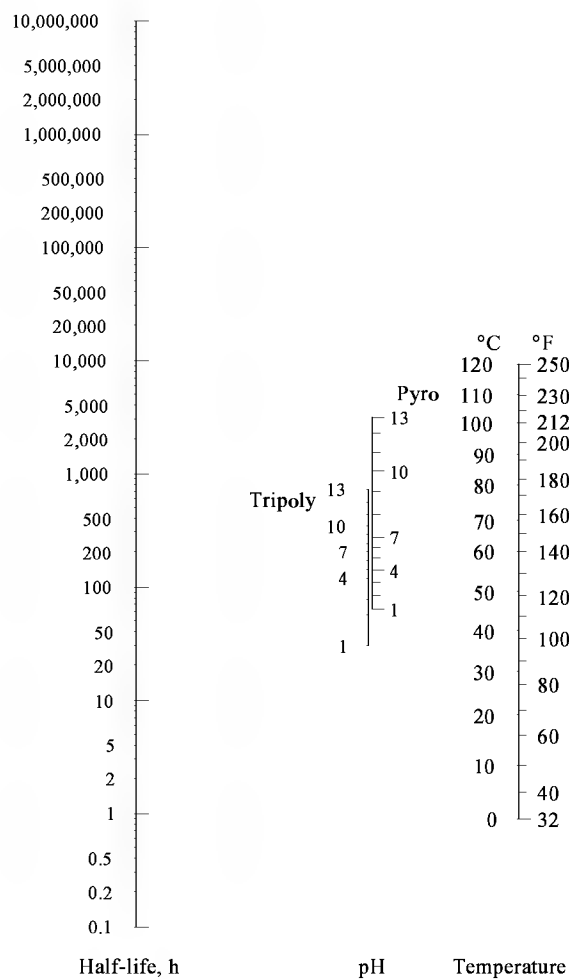


Fig. 10. Nomograph for estimating the rate of hydrolytic degradation of pyrophosphate and tripolyphosphate (tetramethylammonium salts) (27). For sodium salts at pH < 11, multiply half-life by 0.7.

Courtesy of the American Chemical Society.

Many cations have a catalytic effect on hydrolysis, although generally less than that exhibited by hydrogen ions. Hydrolysis rates as a function of pH may exhibit a minimum having higher rates occurring at low pH (H_3O^+ catalysis) and high pH (catalysis by counterion of base, eg, Na^+). For example, for sodium tripolyphosphate, the minimum hydrolysis rate occurs near pH 10, the naturally occurring pH of an STP solution, if NaOH is used to reach pH values higher than 10. The catalytic effect of cations is roughly related to the cation charge size ratio and the cation concentration. Use of quaternary ammonium hydroxides to increase the pH does not result in increased hydrolysis rates because of poor catalysis by the large, low charged quaternary ammonium cation. Phosphatase enzymes catalyze extremely rapid hydrolysis of polyphosphates, at a rate as much as 10^5 times faster than those without enzyme. The activity of these enzymes is highly influenced by a number of factors, including pH and metal ions.

Long-chain polyphosphate hydrolysis is more complex than that of the shorter chains on account of additional mechanistic pathways and the accompanying formation of cyclic metaphosphates. Three mechanisms for hydrolysis of polyphosphate chains in solution are generally recognized: clipping of a monomeric unit from the end of the chain, concurrent loss of three units from the end of the chain by splitting off a trimetaphosphate ion from the end or interior of the chain, and random cleavage from within the interior of the chain to afford shorter chains. All three may occur simultaneously in longer chain polyphosphates, but the last mechanism occurs relatively less frequently. The end group clipping is generally the most prevalent mechanism, except for low pH where random cleavage predominates and in an alkaline environment where trimetaphosphate formation increases in importance. As well as overall degradation rates, higher temperatures increase the proportion of trimetaphosphate to orthophosphate. The overall hydrolysis rate increases with increasing chain length, approaching a limiting value when the average chain length is above 10. In concentrated solutions or with moist solids, the course of the reaction may be more complex. An apparent variation is the dehydration of solid sodium tripolyphosphate hexahydrate, which appears to follow second-order kinetics and initially generates primarily pyrophosphate but little orthophosphate.

Although reasonably stable at room temperature under neutral conditions, tri- and tetrametaphosphate ions readily hydrolyze in strongly acidic or basic solution via polyphosphate intermediates. The hydrolysis is first-order under constant pH. Small cyclic phosphates, in particular trimetaphosphate, undergo hydrolysis via nucleophilic attack by hydroxide ion to yield tripolyphosphate. The ring strain also makes these structures susceptible to nucleophilic ring opening by other nucleophiles.

Complex Ion Formation. Phosphates form water-soluble complex ions with metallic cations, a phenomenon commonly called sequestration. In contrast to many complexing agents, polyphosphates are nonspecific and form soluble, charged complexes with virtually all metallic cations. Alkali metals are weakly complexed, but alkaline-earth and transition metals form more strongly associated complexes (eg, eq. 16). Quaternary ammonium ions are complexed little if at all because of their low charge density. The amount of metal ion that can be sequestered by polyphosphates generally increases

with increasing chain length. For the following reaction, *K* is 10^{5.2}.



Sequestration forms the basis for detergent and water-treatment applications of polyphosphates. Sequestration of hardness ions by sodium tripolyphosphate used in detergent formulations prevents the precipitation of surfactants by the hardness ions. Sodium polyphosphate glass (SHMP) may be added to water system to prevent the formation of calcium or magnesium scales by reducing the activity of the hardness ions. However, if the ratio of cation to polyphosphate is too high at a given pH, insoluble precipitates such as Ca₂P₂O₇ may result instead of the soluble polyphosphate complexes. The application of polyphosphates may be limited in some systems by hydrolytic degradation, depending on the temperatures, pH, and times involved.

Weaker complex ions may also form with orthophosphates, eg, CaH₂PO⁺₄ in mildly acidic solutions of calcium phosphates, and FeHPO⁺₄ as a colorless species in impure phosphoric acid. Anionic complexes such as Fe(HPO₄)₂ are also known.

Colloidal Properties. Polyphosphates are strongly sorbed onto a variety of surfaces where they alter the charge. As a result, the properties of colloidal systems can be dramatically changed by the addition of small amounts of polyphosphate. A striking example is the deflocculation of clays. If a hundredth of a percent of sodium polyphosphate glass is worked into a firm lump of kaolin, the entire mass liquefies to a thin slurry, thereby allowing the clay to be pumped. The same principle of altering the surface charge to repel particles can be used to suspend inorganic materials in aqueous solution.

Calcium carbonate (calcite) scale formation in hard water can be prevented by the addition of a small amount of soluble polyphosphate in a process known as threshold treatment. The polyphosphate sorbs to the face of the calcite nuclei and further growth is blocked. Polyphosphates can also inhibit the corrosion of metals by the sorption of the phosphate onto a thin calcite film that deposits onto the metal surface. When the polyphosphate is present, a protective anodic polarization results.

Analysis. Excellent reviews of phosphate analysis are available (28). Solids characterization methods such as x-ray powder diffraction (xrd) and thermal gravimetric analysis (tga) are used for the identification of individual crystalline phosphates, either alone or in mixtures. These techniques, along with elemental analysis and phosphate species determination, are used to identify unknown phosphates and their mixtures. Particle size analysis, surface area, microscopy, and other standard solids characterizations are useful in relating solids properties to performance. Solid-state nmr is used with increasing frequency.

In most analytical procedures for determining the total phosphorus content (normally expressed in terms of P₂O₅), the phosphates are converted to the orthophosphate form. Typically, condensed phosphates are hydrolyzed to orthophosphate by boiling in dilute mineral acid (0.1 N). The orthophosphate is then determined by gravimetric or spectrophotometric methods. For gravimetric determination, insoluble phosphomolybdates (or magnesium ammonium orthophosphate) is formed.

For the determination of condensed phosphate species, hydrolytic conditions must be avoided. Many of the wet analytical methods for species identification, such as selective precipitation, paper, and thin-layer chromatography, have been largely displaced by spectroscopic or automated methods. Advances have made techniques such as ion chromatography, ³¹P nuclear magnetic resonance, and isotachopheresis practical methods for quantitative analysis. Ion chromatography and ion-exchange chromatography, employing strong-base resins, are widely used for phosphate species analysis. Gradient elution is employed, and excellent resolution of linear phosphates up to 14–18 phosphorus atoms can be obtained. Spectrophotometric analysis following ion-exchange separation and hydrolysis is commonly automated.

MANUFACTURE OF PHOSPHATE SALTS

The general manufacturing scheme for phosphate salts is shown in Figure 11. Condensed phosphates are prepared from the appropriate orthophosphate or mixture of orthophosphates, so the preparation of orthophosphates must be considered first for the manufacture of any phosphate salt. Phosphoric acid is neutralized to form a solution or slurry with a carefully adjusted acid base ratio according to the desired orthophosphate product. The orthophosphate may be recovered either by crystallization from solution, or the entire solution or slurry may be evaporated to dryness. The dewatering (qv) method is determined by the solubility properties of the product and by its desired physical properties such as crystal size and shape, bulk density, and surface area. Acid orthophosphate salts may be converted to condensed phosphates by thermal dehydration (calcination).

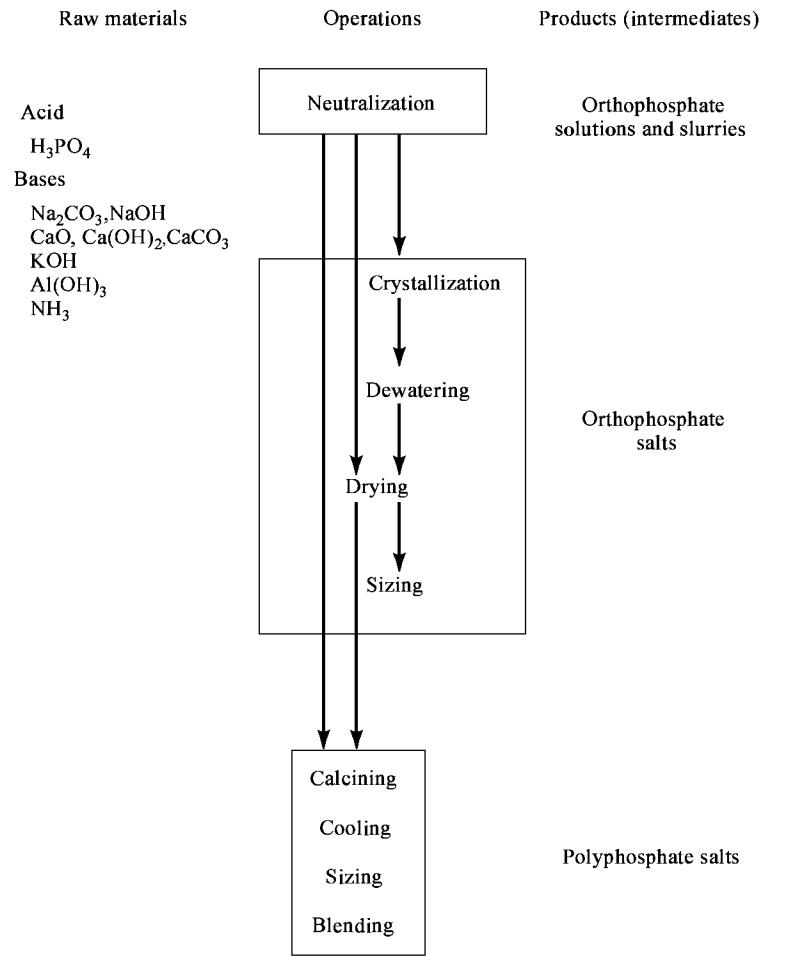


Fig. 11. General manufacturing scheme for phosphate salts.

Orthophosphates.

Alkali Metal Phosphates. A significant proportion of the phosphoric acid consumed in the manufacture of industrial, food, and pharmaceutical phosphates in the United States is used for the production of sodium salts. Alkali metal orthophosphates generally exhibit congruent solubility and are therefore usually manufactured by either crystallization from solution or drying of the entire reaction mass. Alkaline-earth and other phosphate salts of polyvalent cations typically exhibit incongruent solubility and are prepared either by precipitation from solution having a metal oxide : P_2O_5 ratio considerably lower than that of the product, or by drying a solution or slurry with the proper metal oxide : P_2O_5 ratio.

Monosodium phosphate, NaH_2PO_4 , is usually marketed as the anhydrous salt made by either crystallization from solution or evaporation of a solution to dryness on steam-heated drum dryers and followed by drying in a rotary dryer. Disodium phosphate, Na_2HPO_4 , is produced commercially as the anhydrous salt and as the dihydrate, $Na_2HPO_4 \cdot 2H_2O$. The dihydrate is crystallized from solution between 40–95°C, followed by centrifugation and drying. Anhydrous Na_2HPO_4 , precipitated from solution above 95°C, exhibits inverse solubility and may be crystallized by boiling off water (see Fig. 7). The anhydrous salt may also be obtained by dehydration of the dihydrate in a rotary dryer.

Trisodium phosphate (TSP) is manufactured as the anhydrous salt, Na_3PO_4 , and as an incompletely dehydrated salt (with a water content equivalent to a monohydrate) by drying a solution or slurry above 80°C. Spray drying, drum drying, and flash drying are used for the evaporation step, but rather high (300–400°C) drying temperatures are needed for complete dehydration. Both salts dissolve incongruently in water, affording the dodecahydrate (crystalline TSP) with an $Na_2O : P_2O_5$ ratio of about 3.25.

Crystalline TSP is a dodecahydrate with somewhat variable composition between the limits of $(Na_3PO_4 \cdot 12H_2O) \cdot 0.25NaOH$ and $(Na_3PO_4 \cdot 12H_2O) \cdot 1.7NaOH$. It is manufactured by crystallization below 60°C from a solution with an $Na_2O : P_2O_5$ mole ratio slightly lower than 3.25. Crystals are isolated by centrifugation and air-dried at ca 40°C to minimize dehydration.

Chlorinated TSP (Cl-TSP) is a complex mixture or solid solution approximating the composition $4(Na_3PO_4 \cdot 11H_2O) \cdot NaOCl$. The composition is variable, having an $Na_2O : P_2O_5$ mole ratio ranging from ca 3.15–3.35, and 3–5% available chlorine. Sodium chloride and $Na_2HPO_4 \cdot 2H_2O$ are present as impurities in the commercial product. It is manufactured by the addition of sodium hypochlorite solution to a hot, concentrated sodium phosphate solution having an $Na_2O : P_2O_5$ mole ratio of ca 2.85, followed first by cooling, crystallization, and granulation of the entire reaction mass in a heavy-duty mixer, then by low temperature air-cooling and drying. Chlorinated TSP is unstable above 40°C and up to 20% of the available chlorine in the initial reaction mass may be lost. Chlorine stability is improved by the addition of 0.5% sodium silicate to the reaction mass, cooling the mass rapidly during granulation, applying refrigerated air in the final drying step, and protecting against high ambient temperatures during storage and transportation.

Potassium Phosphates. Potassium phosphate salts are analogous to the sodium salts and share many of the same functional properties. The higher cost of potassium hydroxide has restricted these salts to applications where high solubility or nutrient value is important. Potassium salts are manufactured like their sodium analogues, often on the same equipment. Many of the potassium phosphates are more deliquescent than their sodium analogues and may require special storage and moistureproof containers.

Ammonium Phosphates. In the manufacture of ammonium phosphates, an atmosphere of ammonia may need to be maintained because the partial pressure of ammonia rises rapidly as either the temperature or the $NH_3 : P_2O_5$ mole ratio of the reaction mass increases. Phosphoric acid reacts quickly with ammonia vapor and is used in multistage reactor systems as a scrubber fluid to prevent NH_3 emissions and recover ammonia values. For example, H_3PO_4 scrubbing of coke-oven off-gases produces ammonium phosphates of relatively good purity.

Monoammonium phosphate (MAP), $NH_4H_2PO_4$, is produced by reaction of anhydrous ammonia and phosphoric acid in batch or continuous reactors and crystallized in conventional crystallizers because the partial pressure of ammonia over this acidic solution is relatively low (see CRYSTALLIZATION). Crystals are centrifuged and dried below 100°C in a rotary dryer, and mother liquor is returned to the reactor. Diammonium phosphate (DAP), $(NH_4)_2HPO_4$, solutions, on the other hand, have a high partial pressure of ammonia and the reaction is carried out in a two-stage reactor system in which the feed acid passes countercurrentwise to the flow of ammonia gas. Incoming acid reacts in the scrubber with ammonia from the main reactor and may also serve as a scrubber for dryer off-gases. MAP and DAP fertilizers are made in a granulation process from ammonia and wet-process phosphoric acid. The acid is partially neutralized in a tank reactor, and ammoniation and granulation are completed in a rotary drum. Drying, cooling, and product screening complete the process.

Calcium Phosphates. Because calcium phosphates and calcium bases as raw materials have low solubilities, the manufacture of calcium phosphates must therefore deal with nonequilibrium chemistry in a heterogenous system. Furthermore, several kinetically favored metastable phosphates occur under manufacturing conditions and these salts may hydrolyze to the highly insoluble and stable hydroxyapatite. Calcium phosphates also exhibit incongruent solubility, which means that the mother liquor necessarily has a different $CaO : P_2O_5$ ratio than the product (Fig. 12). As a result, most commercial calcium phosphates are mixtures of several salts having an average composition approximating that of the pure material. For example, the addition of a slurry of hydrated lime into a well-stirred phosphoric acid solution until a pH of 7 is reached results in the precipitation of dicalcium phosphate, which can be anhydrous or dihydrate, depending on the temperature. If, however, the order of addition is reversed and the phosphoric acid is added into a well-stirred lime slurry until a pH of 7 is reached, tricalcium phosphate (TCP) is obtained. Each product can contain up to about 10% of the other. The TCP product is actually an amorphous basic calcium phosphate similar to hydroxyapatite, $Ca_{10}(PO_4)_6(OH)_2$, rather than $Ca_3(PO_4)_2$, as indicated by the name tricalcium phosphate.

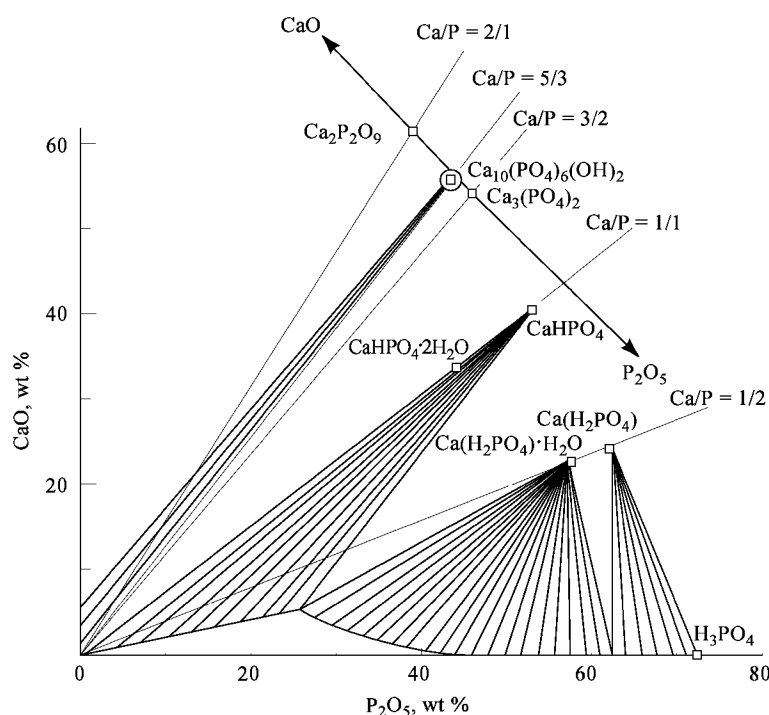


Fig. 12. Phase diagram of the CaO–H₂O–P₂O₅ (calcium orthophosphate) system where the circle represents the variable hydroxylapatite composition and (□) the molecular species indicated.

For fertilizer and animal nutrition uses, the primary concern is the CaO and P₂O₅ analysis of the product, which is usually a mixture of salts. Animal feed-grade dicalcium phosphate is made from wet-process acid that has been treated with superheated steam, or with finely divided silica and steam, to reduce the fluorine content to an acceptable level. Various granulation processes produce granules smaller than those used for fertilizer. For industrial, dentifrice, food, and pharmaceutical uses, important functional properties relating to composition or solids characteristics may significantly differ in two products having nearly identical CaO and P₂O₅ content. Manufacture is something of an art and details are often proprietary.

Monocalcium Phosphates. Monocalcium phosphate (MCP) is generally made as a composition equivalent to the monohydrate, Ca(H₂PO₄)₂·H₂O. The monohydrate is manufactured by several methods. Phosphoric acid and hydrated lime may be mixed in a pan or other heavy-duty mixer that allows the rapid escape of steam. The product is a paste that is dried and sized. The reaction may also be carried out in a more dilute system in conventional mixing equipment to produce a pumpable slurry that is then spray dried to give a lighter and more rapidly soluble product. Both methods, typically running at a CaO : P₂O₅ mole ratio just above 1.0, result in the presence of free acid and dicalcium phosphate as well as unreacted lime particles. Free acid causes caking and MCP is often aged or cured to allow for more complete reaction of the acid. Commercial MCP monohydrate usually contains several percent of dicalcium phosphate, which is acceptable in baking powder as an anticaking agent. MCP may also be manufactured by crystallization from solution. MCP made by this method may contain a small amount of free acid from entrained mother liquor. Thorough washing of an MCP filter or centrifuge cake with water results in partial conversion to dicalcium phosphate.

Anhydrous monocalcium phosphate, Ca(H₂PO₄)₂, can be made in a pan mixer from concentrated phosphoric acid and lime. The high heat of reaction furnishes essentially all the necessary thermal input and subsequent drying is minimized. A small amount of aluminum phosphate or a mixture of sodium and potassium phosphates is added in the form of proprietary stabilizers for coating the particles. Heat treatment converts the coating to a protective polyphosphate (19).

Dicalcium Phosphates. Whereas mixtures of calcium phosphates are made by the pan-mixer process and sold as dicalcium phosphate (DCP) for use as animal feed supplements, the dentifrice-abrasive market requires close control of the product. Formation of minor amounts of more basic calcium phosphates results in rendering soluble fluoride in the dentifrice formulation as an inactive and insoluble fluorapatite. A much more dilute slurry reaction is used to obtain good mixing. The more abrasive anhydrous salt, CaHPO₄, is precipitated at ca 80°C, whereas the softer dihydrate precipitates below 45°C. Conventional mixing equipment is used for the reaction followed by centrifugation, drying, and milling to the desired particle size. Stability of the DCP dihydrate (DCPD) against reaction with fluoride in toothpaste formulations is of tantamount importance. Fluoride, typically added in 0.1% levels, reacts with phosphates more basic than DCPD to form fluorapatite, thereby inactivating the caries-preventative action of the free fluoride ion. For this reason, the presence of calcium phosphate impurities can be tolerated at only low levels. The dihydrate is often stabilized against disproportionation to more basic calcium phosphates by sodium pyrophosphate and/or trimagnesium phosphate.

Tricalcium Phosphate. Commercial tricalcium phosphate (TCP) is actually an amorphous basic calcium phosphate close to hydroxyapatite in composition. Because of its extremely low solubility in water, TCP is precipitated almost quantitatively from dilute phosphate solutions with a slurry of hydrated lime. TCP is separated by drum-, spray-, or flash-drying the TCP slurry, with or without intermediate sedimentation or filtration steps. It is used as an industrial-grade flow conditioner and parting agent.

Condensed Phosphates. Condensed phosphates are prepared by calcining an orthophosphate composition having the proper metal oxide : P₂O₅ mole ratio. Depending on the metal oxide : P₂O₅ mole ratio of the desired product, this calciner feed composition may include compound(s) of a single metal oxide : P₂O₅ mole ratio or a mixture of compounds having the proper overall ratio. Commercial practice for the manufacture of the condensed sodium or potassium phosphates typically begins with the preparation of an orthophosphate feed liquor (solution) having the proper M₂O : P₂O₅ ratio. The liquor may be processed in two stages by using separate dryer and calciner, or by using specially designed combination dryer-calciner units (29) such as those shown in Figure 13, in which liquor may be sprayed directly onto a hot recycled bed of calcined material. The drying conditions determine in large measure the physical properties of the calcined product, such as bulk density. Rotary calciners, either gas- or oil-fired, are preferred. Following calcination, the product is water-cooled in screw or tube coolers and then sized to granular or powder specifications by screening, air separation, and milling techniques.

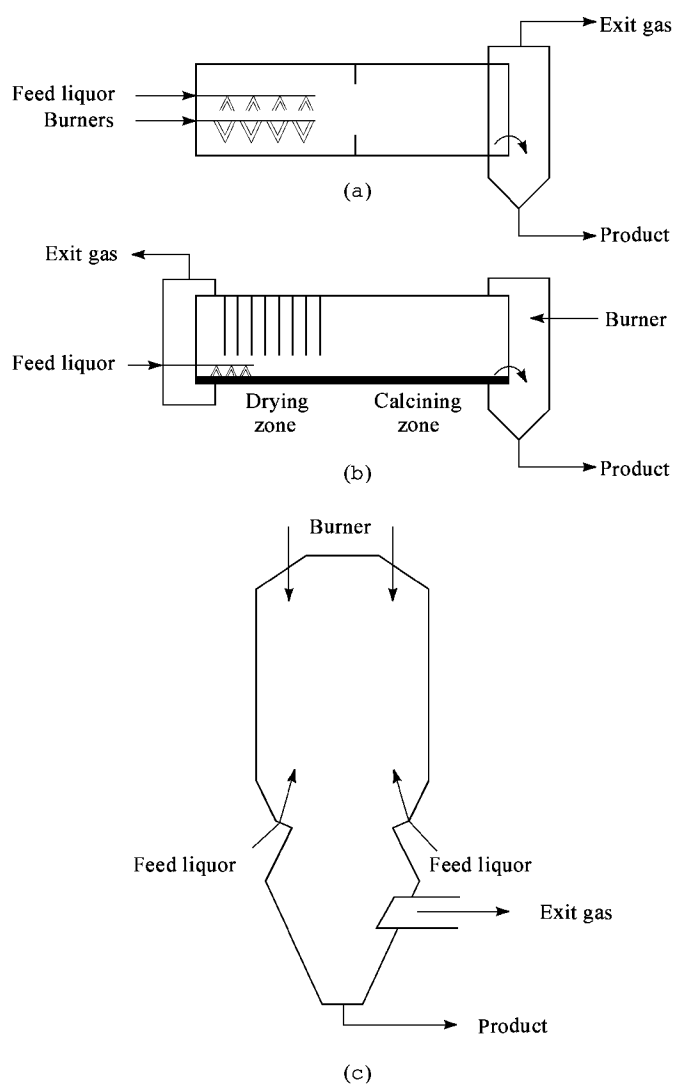


Fig. 13. Combined dryer-calciner processes for sodium tripolyphosphate: (a) cocurrent rotary dryer-calciner; (b) countercurrent rotary dryer-calciner; and (c) spray dryer-calciner.

Three crystalline polyphosphates can be produced by thermal dehydration of monosodium phosphate at successively higher temperatures (see Fig. 9). Monosodium phosphate (MSP) may be used in any form that has a $\text{Na}_2\text{O} : \text{P}_2\text{O}_5$ mole ratio of 1.0, such as NaH_2PO_4 , $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$, or a slurry or solution of MSP. When heated to about 250°C , MSP converts to sodium acid pyrophosphate (SAPP). For leavening applications, SAPP is manufactured to exacting specifications on the rate of reaction in dough mixtures. Many grades are produced for different leavening applications by various calcination procedures and the addition of trace quantities of other cations and phosphates. Slow and controlled calcination of MSP or SAPP to about 400°C produces insoluble metaphosphate (IMP), a long-chain, crystalline sodium polyphosphate of the Maddrell's salt type. Sodium trimetaphosphate (STMP), $\text{Na}_3\text{P}_3\text{O}_9$, is produced by calcination of MSP, SAPP, STMP, or IMP at $450\text{--}500^\circ\text{C}$. The rates at which the thermal dehydration reactions and crystalline transitions occur depend on temperature profiles, water vapor pressures, impurities, etc. As a consequence, commercial polyphosphates usually contain several percent of one or more of the other members of this group.

Sodium polyphosphate compositions corresponding to average chain lengths from four up to, but excluding, very long chains are amorphous glassy mixtures of various chain lengths, as opposed to crystalline compositions. These sodium phosphate glasses are generally prepared having $\text{Na}_2\text{O} : \text{P}_2\text{O}_5$ mole ratios of approximately 1.1–1.3, and average chain lengths of about 21–27 monomeric phosphate units, respectively. Although the glasses can be made at any molar ratio up to about 1.7, many of the solution properties are similar to that of the crystalline sodium tripolyphosphate. Glass nomenclature is inexact and many trade names are used for glasses of slightly different composition. Glasses are manufactured in refractory-type reverberatory furnaces in which a solution (or less typically a solid feed) is brought to a melt approaching 1000°C and then rapidly quenched on water-cooled stainless steel belts, rolls, or wheels. These processes are energy- and maintenance-intensive.

Sodium tripolyphosphate, $\text{Na}_5\text{P}_3\text{O}_{10}$, is manufactured from a solution having $\text{Na}_2\text{O} : \text{P}_2\text{O}_5$ mole ratios near 1.67, corresponding to a mixture of two moles of disodium phosphate and one mole of monosodium phosphate. Mixtures of phosphate and pyrophosphate salts can also be used, provided that the 1.67 molar ratio is maintained. Because the precursor phosphate salts can convert upon calcination to one or more polyphosphates other than sodium tripolyphosphate, the feed must be an intimate mixture of finely divided crystals to ensure a high assay product. For this reason, most commercial processes start with a carefully adjusted, homogeneous solution that is then dried as rapidly as possible to minimize individual crystal growth. STP with a tripolyphosphate assay as high as 98–99% can be produced from phosphate feeds dried by spray dryer, or in rotary dryers in which the solution is flash-dried on a hot, rolling bed of material. Drum and rotary dryers that evaporate the feed solution more slowly and allow better crystal growth, however, lead to STP having an assay usually no higher than 95%. Some older process dryers allow growth of large crystals that cannot be converted to tripolyphosphate of an assay over ca 85%. However, higher conversion efficiency can be obtained by milling the dried phosphate, calcination in the presence of a high partial pressure of water, and post-treatment of hot calciner product in a steam atmosphere (30). Physical properties of the STP, such as bulk density, dissolution rate, etc, depend in large measure on the drying method. Aside from considerations of assay, the drying method is often selected for control of the STP physical properties. Calcining is typically carried out at $380\text{--}500^\circ\text{C}$, depending on the proportion of STP-I and -II desired.

In most condensed phosphate manufacturing operations using rotary calciners, a mixture of powder and granular product is formed. As a result of newer methods of detergent manufacture (eg, agglomeration and dry mix, as opposed to spray tower methods) and the movement away from phosphate use in home laundry detergents in general, the U.S. STP market has steadily shifted from powder to granular products. Three granular bulk-density-grades are recognized: light at $\sim 0.5 \text{ g cm}^{-3}$, medium at $\sim 0.75 \text{ g cm}^{-3}$, and heavy at $\sim 1.0 \text{ g cm}^{-3}$. In most processes, the granular bulk density is fixed at the phosphate drying stage; thus spray dryers produce light-, drum dryers medium-, and rotary dryers heavy-density granules. Bulk density can be controlled over a wide range in one process by manipulating particle size and moisture content of the feed to an agglomerating calciner (31). Other processes have been developed to convert excess powder or by-product fines from granular operations to a granular product by reagglomeration with water or a

phosphate solution and recalcination (32).

Tetrasodium pyrophosphate (TSPP), Na₄P₂O₇, is obtained by the calcination of disodium phosphate, or any of its hydrates, at about 400°C.

Commercial manufacture is similar to that of sodium tripolyphosphate, often using the same equipment. Relatively little calcining control is required for TSPP because it is the only condensation product of disodium phosphate (DSP). Owing to increased dust generation, however, pyrophosphate processing typically requires some changes in dust-handling procedures.

Ammonium Polyphosphate. The poorly soluble, long-chain crystalline ammonium polyphosphate (NH₄PO₃)_n, is used in intumescent fire-retardant coatings and paints where resistance to leaching by water is required. It is manufactured by heating a mixture of urea (qv) and ammonium phosphate or polyphosphoric acid under a controlled atmosphere of NH₃ and water at ~300°C (33). Urea acts as a condensing agent because the polyphosphate cannot be made by simple thermal dehydration on account of the high partial pressure of ammonia. The higher temperature Form II ammonium polyphosphate is used as a fire retardant in thermoplastics.

Calcium Pyrophosphate. Calcium pyrophosphate, Ca₂P₂O₇, is manufactured by high temperature calcination of DCP in a rotary calciner.

Temperature is carefully controlled to adjust the proper ratio of β- and γ-forms.

Economic Aspects

The estimated world production of wet-process phosphoric acid was 24,001,000 metric tons of P₂O₅ in 1993. Capacity was 34,710,000 metric tons. Over 90% of phosphoric acid production is wet-process (agricultural-grade) acid; the remainder is industrial-grades (technical, food, pharmaceutical, etc) made by the thermal route or by the purification of wet-process acid. Table 11 lists U.S. production of wet-process and industrial-grade acids.

Table 11. U.S. Production of Phosphoric Acid^a, t × 10³/yr

Year	Industrial-grade ^b	Wet-process agricultural-grade
1985	685	7,544
1986	696	6,663
1987	699	7,593
1988	722	8,556
1989	700	8,671
1990	694	8,990
1991	636	9,300
1992	555	10,179
1993	570	8,532

^a Ref. 34.

^b Technical- and food-grade.

U.S. consumption of industrial-grade phosphoric acid and phosphates in 1993 according to product categories (34) was phosphoric acid, at 29% ; sodium phosphate, 52% ; calcium phosphate, 7% ; potassium phosphate, 3% ; ammonium phosphate, 5% ; and others, 4% . Consumption according to market is given in Table 12.

Table 12. U.S. Consumption of Industrial-Grade Phosphoric Acid and Phosphate Salts in 1989^a

Market	Acid, %	Salts, %
builders and water treatment	15	70
foods	13	17
metals	17	<1
others	55	13
<i>Total</i>	<i>100</i>	<i>100</i>

^a Ref. 34.

The production of industrial-grade acid and phosphates has been influenced since the 1970s by environmental concerns and related legislation. Industry consolidation and rationalization have resulted. A shift in production toward wet-acid purification routes has occurred and only the most economically viable elemental phosphorus and thermal acid producers remain in business. There has also been an increased focus on higher value phosphate products as STP volumes have declined. Prices of phosphoric acids and selected industrial-grade phosphates are listed in Table 13.

Table 13. Prices for Phosphoric Acid and Selected Phosphates^a in 1995

Material	Price, \$ /kg		
	Technical-grade	Food-grade	Agricultural-grade
phosphoric acid, % H ₃ PO ₄			
53 ^b			298 ^c
70 ^b			336 ^c
75	0.6724	0.7165	
80	0.7055	0.7496	
85	0.7496	0.7937	
monoammonium phosphate	1.301	1.378	
diammonium phosphate	1.317	1.345	
monocalcium phosphate		1.240	
dicalcium phosphate		1.405 ^d	
sodium tripolyphosphate	0.942	1.146	

^a Ref. 35.

^b Value is % P₂O₅.

^c Price is per metric ton P₂O₅.

^d Dentifrice-grade.

Safety and Environmental Considerations

Inorganic phosphates present little hazard to humans and are mineral nutrients essential to life processes. Attention must be given to the acidity of phosphoric acid, the alkalinity of the bases with which it reacts and the heat released upon neutralization. Appropriate protective gear should be worn when in close contact. Some phosphate salts are reasonably acidic or basic.

Larger environmental issues are associated with the manufacture of wet-process acid and elemental phosphorus, than with the manufacture of technical- or food-grade acids and salts from these raw materials. In the manufacture of both wet acid and phosphorus, the P₂O₅ value recovered may represent a minor proportion of the phosphate rock as mined. With the exception of the removal of traces of arsenic, phosphorus can be cleanly converted to thermal acid and then to salts. Purified wet acid is also converted into salts with little waste generation. The raffinate and by-product streams from wet-acid purification are generally returned to the fertilizer complex from where the wet-acid feed originated.

Because of the nutritive value, phosphates have been implicated in promoting the growth of algae in lakes. Problems apparently caused by sewage-borne phosphates are mostly localized to areas that have traditionally employed lakes as receiving waters for sewage effluents. It is believed that much of the phosphate is precipitated in an insoluble form and trapped in sediments where it is ultimately converted to an apatite. Considerable controversy has centered on the contribution of phosphate-built detergents to excessive algae growth and subsequent eutrophication of natural receiving water. Legislation against the use of phosphates in detergents has resulted in a patchwork of restrictions worldwide. Home laundry detergents have been the most regulated. Societal pressure has resulted in the voluntary reduction or elimination of phosphates in many cleaning products by the manufacturers. It is open to question, however, as to whether a banning of phosphate detergents and cleaners can indeed sufficiently reduce phosphorus input to the low levels needed to control algal growth, when, in fact, natural wastes and fertilizers provide most of the phosphorus input to receiving waters. A more logical but also more costly approach is phosphorus removal during sewage treatment. Excellent reviews of this area are available (36,37).

BIBLIOGRAPHY

"Phosphoric Acids and Phosphates" in *ECT* 1st ed., Vol. 10, pp. 403–411, J. R. Van Wazer, Monsanto Chemical Co.; in *ECT* 2nd ed., Vol. 15, pp. 234–276, by J. R. Van Wazer, Monsanto Co.; in *ECT* 3rd ed., Vol. 17, pp. 426–472, by R. B. Hudson and M. J. Dolan, Monsanto Co.

1. J. E. Such, in A. A. Eldridge, G. M. Dyson, A. J. E. Welch, and D. A. Pantony, eds., *Mellor's Comprehensive Treatise on Inorganic and Theoretical Chemistry*, Longman, London, U.K., Vol. VIII, Suppl. III, 1971, p. 726.
2. U.S. Pat. **2,708,619** (May 17, 1955), H. S. Winnicki and C. P. Roberts (to FMC Corp.).
3. H. Y. Allgood and co-workers, *Ind. Eng. Chem.* **59**(6), 18 (1967).
4. J. A. Brink, Jr., *Can. J. Chem. Eng.* **41**, 134 (1963).
5. U.S. Pat. **3,129,057** (Apr. 14, 1964), L. Germain and P. Desire (to Progil S.A.).
6. Eur. Pat. **0 037 735 B1** (July 31, 1985), R. B. Hudson, G. W. Cadwallader, and D. C. Taylor (to Monsanto Co.).
7. Jpn. Kokai **79-84890** (July 6, 1979), Y. Takahashi and T. Kobayashi (to Kawasaki Heavy Ind. Ltd.).
8. E. J. Lowe, *Proceedings of the Second International Congress on Phosphorus Compounds*, 1980, p. 493.
9. H. M. Stevens, in J. R. Van Wazer, ed., *Phosphorus and Its Compounds*, Vol. 2, Interscience Publishers, New York, 1961.
10. U.S. Pat. **3,914,382** (Oct. 21, 1975), T. A. Williams (to Albright and Wilson, Ltd.).
11. R. Fitoussi and J. Helgorsky, *Ind. Miner., [Ser.] Tech.* (3) 225 (1982).
12. A. Davister and M. Peeterbroeck, *Chem. Eng. Prog.* **78**(3), 35 (1982).
13. Ger. Offen. **2,334,019** (Jan. 30, 1975), B. Wojtech, K. P. Ehlers, and W. Scheibitz (to Farbwerke Hoechst A.G.).
14. R. F. Jameson, *J. Chem. Soc.*, 752 (1959).
15. E. H. Brown and C. D. Whitt, *Ind. Eng. Chem.* **44**, 615 (1952).
16. J. R. Van Wazer, ed., *Phosphorus and Its Compounds*, Vol. 1, Interscience Publishers, New York, 1958, p. 493.
17. B. Wendrow and K. A. Kobe, *Chem. Rev.* **54**, 891 (1954).
18. D. R. Gard, in J. J. McKetta and W. A. Cunningham, eds., *Encyclopedia of Chemical Processing and Design*, Marcel Dekker, Inc., New York, 1990, Vol. 35, pp. 429–495.
19. U.S. Pat. **2,462,104** (Feb. 22, 1949), W. H. Knox and co-workers (to Victor Chemical Co.); U.S. Pat. **2,631,102** (Mar. 10, 1953), F. E. Hubbard and J. S. Metcalf (to Monsanto Co.).
20. T. Kanazawa, ed., *Inorganic Phosphate Materials*, Kodansha, Tokyo, Japan, 1989.
21. U.S. Pat. **4,515,772** (May 7, 1985), J. J. Parran, Jr. and N. Y. Sakkab (to Procter and Gamble Co.); U.S. Pat. **4,627,977** (Dec. 9, 1986), A. Gaffar and T. G. Polefka (to Colgate-Palmolive Co.).
22. J. G. Rabatin, R. H. Gale, and A. E. Newkirk, *J. Phys. Chem.* **64**, 491 (1960).
23. T. D. Farr, *Phosphorus, Properties of the Element and Some of Its Compounds*, in *Chemical Engineering Report*, No. 8, Tennessee Valley Authority, Wilson Dam, Ala., 1950, pp. 39 and 58.
24. J. D. McGilvery, *ASTM Bull.* (19), 45 (1933).
25. C. Y. Shen and R. A. Herrmann, *Ind. Eng. Chem. Prod. Res. Dev.* **5**(4), 357 (1966); A. C. Zettlemoyer and C. H. Schneider, *J. Am. Chem. Soc.* **78**, 3870 (1956).
26. C. Y. Shen and C. F. Callis, *Soap Chem. Specialities*, **41**(2), 64 (1965); C. Y. Shen, *Ind. Eng. Chem., Prod. Res. Develop.* **5**(3), 272 (1966).
27. E. J. Griffith, *Ind. Eng. Chem.* **51**, 240 (1959).
28. M. Halmann, ed., *Analytical Chemistry of Phosphorus Compounds*, Wiley-Interscience, New York, 1972.
29. U.S. Pat. **3,661,514** (May 9, 1972), J. Herink (to FMC Corp.); U.S. Pat. **2,419,148** (Apr. 15, 1947), C. S. King (to Blockson Chemical Co.).
30. U.S. Pat. **3,309,174** (Mar. 14, 1967), R. Pals (to FMC Corp.).
31. U.S. Pat. **3,233,967** (Feb. 8, 1966), C. Y. Shen (to Monsanto Co.).
32. U.S. Pat. **3,932,590** (Jan. 13, 1976), R. Pals, R. J. Fuchs, and J. C. Schwartz (to FMC Corp.); U.S. Pat. **4,251,498** (Feb. 17, 1981), P. Hensler, D. F. Church, and R. W. Beger (to FMC Corp.); U.S. Pat. **4,255,274** (Mar. 10, 1981), P. Hensler, D. F. Church, and R. W. Beger (to FMC Corp.).
33. U.S. Pat. **3,495,937** (Feb. 17, 1970), C. Y. Shen (to Monsanto Co.).
34. M. Smart, *Industrial Phosphates*, Chemical Economics Handbook–SRI International, Menlo Park, Calif., 1991; D. H. Lauriente, *Phosphate Industry Overview*, Chemical Economics Handbook–SRI International, Menlo Park, Calif., 1994.
35. H. Van, ed., *Chem. Mark. Rep.* **246**(7) (Aug. 15, 1994).
36. E. J. Griffith and co-workers, eds., *Environmental Phosphorus Handbook*, Wiley-Interscience, New York (1973).
37. *Phosphorus in the Environment: Its Chemistry and Biochemistry*, Ciba Foundation Symposium 57, Excerpta Medica, Amsterdam, 1979.

General References

J. R. Van Wazer, ed., *Phosphorus and Its Compounds*, Vol. 1 (1958) and Vol. 2 (1961), Interscience Publishers, Inc., New York.
D. E. C. Corbridge, *Phosphorus: An Outline of Its Chemistry, Biochemistry, and Technology*, 4th ed., Elsevier, Amsterdam, the Netherlands, 1990.
K. Schrüdter and co-workers, in B. Elvers, S. Hawkins, and G. Schulz, eds., *Ullmann's Encyclopedia of Industrial Chemistry*, VCH, Weinheim, Germany, 1991, Vol. A19, 465–503.

A. D. F. Toy, *Phosphorus Chemistry in Everyday Living*, American Chemical Society, Washington, D.C., 1987.
P. Becker, *Phosphates and Phosphoric Acid*, 2nd ed., Marcel Dekker, Inc., New York, 1989.

David R. Gard
Monsanto Company

PHOSPHORS.

See LUMINESCENT MATERIALS; PHOTODETECTORS.

PHOSPHORUS

Phosphorus [7723-14-0] is a nonmetallic element having widespread occurrence in nature as phosphate compounds (see PHOSPHORIC ACID AND PHOSPHATES). Fluorapatite [1306-05-4], Ca₅F(PO₄)₃, is the primary mineral in phosphate rock ores from which useful phosphorus compounds (qv) are produced. The recovery from the ore into commercial chemicals is accomplished by two routes: the electric furnace process, which yields elemental phosphorus; and the wet acid process, which generates phosphoric acid. The former is discussed herein (see FURNACES, ELECTRIC). Less than 10% of the phosphate rock mined in the world is processed in electric furnaces. Over 90% is processed by the wet process, used primarily to make fertilizers (qv).

Most of the phosphorus produced as the element is later converted to high purity phosphoric acid and phosphate compounds; the remainder is used in direct chemical synthesis to produce high purity products. In contrast, phosphoric acid produced by the wet process is used in lower purity applications, especially in fertilizer and to a lesser degree in animal feed (see FEEDS AND FEED ADDITIVES). More recently, a small portion of wet acid is purified in a second process and then also used in high purity acid and phosphate compound applications.

Elemental phosphorus is produced and marketed in the α-form of white or yellow phosphorus, the tetrahedral P₄ allotrope. A small amount of red amorphous phosphorus, P, is produced by conversion from white phosphorus. White phosphorus as the element is characterized by its combustion in air to form phosphorus pentoxide. Consequently, white phosphorus is generally stored and handled under water. Elemental white phosphorus is also highly toxic, and suitable precautions are required by those who manufacture or handle it. The black phosphorus modification prepared under high pressure does not have commercial importance.

The history of elemental phosphorus has been reviewed (1). Discovery is credited to Hennig Brandt in 1669. Early commercial production was by reduction of phosphoric acid using carbon in the 1830s and by the reduction of bones in the 1840s. The main application was in matches (qv), which has been discontinued owing to the toxicity of white phosphorus. In 1888 the electric furnace process was patented by Readman in Britain. This process has been developed over the years. A significant increase in U.S. production occurred beginning in the 1930s in Tennessee and Alabama, driven by the low cost of electric power from the Tennessee Valley Authority and the high purity of phosphoric acid made by the electric furnace route as compared to the wet-acid process. A further expansion took place in the 1940s as a result of the development of detergents based on phosphates, especially sodium tripolyphosphate. North American production of elemental phosphorus reached its zenith in the early 1970s. Production has declined since then because of reformulations of detergent phosphates. References 1–7 contain additional information.

Physical Properties

The dominant commercial form of elemental phosphorus is the α-white allotrope. α-White phosphorus is often designated simply as P₄ because the solid consists of tetrahedral P₄ molecules. In its pure form, it is a white solid that forms a clear liquid when melted. However, the commercial product is generally somewhat yellow, both as a solid and as a liquid, owing to the presence of small amounts of a red phosphorus allotrope. Commercial white phosphorus may also be slightly gray in color because of incomplete separation of coke dusts and other impurities generated in the manufacturing process.

Upon storage, the amount of red phosphorus in solid white or liquid phosphorus may increase if the material is exposed to light or contains contaminants such as iodine, sulfur, selenium, or sodium that catalyze the conversion from white to red. Also, because white phosphorus is generally stored under water, some surface oxidation to form viscous white or colored polymeric oxyacids also occurs, especially if the oxygen content of the water can be replenished by exposure to air.

White phosphorus is a soft waxy solid often compared to paraffin wax. Small samples under water can be cut with a knife with some difficulty. Large samples are always transferred and handled as a liquid under water, which protects the phosphorus from air oxidation.

The solubility of phosphorus in water is about 3 ppm. However, process water used in phosphorus manufacture or handling often carries larger amounts of phosphorus as particulates or small droplets, depending on the water temperature. Phosphorus-contaminated water is commonly called phossy water. Phosphorus has low solubility in most common solvents, but is quite soluble in carbon disulfide and some other special solvents. The solubility in CS₂ and benzene was formerly used in phosphorus analyses, but toxicity and increasing waste disposal costs have led to more use of toluene and xylene, and more recently to the use of nonchemical turbidity measurements.

Other phosphorus allotropes can be made from the α-form. At temperatures below –76.9°C a hexagonal β-white phosphorus having a higher (1.88 g cm³) density can be produced reversibly. The difference between the α- and β-forms has been shown by ³¹P nmr to be that the P₄ tetrahedra are not free to rotate in the β-white modification.

Some of the more common physical properties of α-white phosphorus are given in Table 1. White phosphorus is easily melted at low (44°C) temperatures using hot water, facilitating its transfer and handling. Like solid white phosphorus, the liquid has been shown to consist of P₄ tetrahedra. Upon cooling, the liquid P₄ shows a strong tendency to supercool several degrees and then to solidify with a high freezing rate. Cubic, transparent crystals can be crystallized from the melt or from phosphorus vapor. However, most solid phosphorus does not show crystal faces, but rather freezes into the massive waxy solid. Freezing liquid phosphorus yields α-white phosphorus.

Table 1. Physical Properties of α-White Phosphorus, P₄^a

Parameter	Value	
melting point, °C	44.1	
boiling point, °C	280.5	
heat of fusion, J (kgmol) ^b	2.5 × 10 ⁶	
heat of vaporization, J (kgmol) ^b	49.8 × 10 ⁶	
density, g cm ³		
solid	1.83	
liquid at 50°C	1.74	

liquid viscosity, at 50°C, mPas (cP)	1.69
solubility of P ₄ , g 100 g solvent	
carbon disulfide at 10°C	89.8
benzene at 15°C	2.7
water at 15°C	0.00033

^a See especially Ref. 3.

^b To convert J to cal, divide by 4.184.

Upon heating, α -white phosphorus first melts, then either vaporizes or converts to amorphous red phosphorus. The conversion to red P proceeds slowly in one to two days at temperatures slightly below the 280°C boiling point of liquid P₄. The product is amorphous to x-ray diffraction.

Besides the amorphous red P which is a commercial product, there are perhaps five other crystalline or poorly crystalline red modifications that are not produced commercially. These include a triclinic, an orthorhombic, a high temperature form, and two poorly crystalline forms. Red phosphorus varieties are rather stable in air and are of lower reactivity than white phosphorus.

Several allotropes of black phosphorus have also been reported (2). These include one amorphous and three crystalline modifications. At increasing pressures and temperatures reaching above 1200 MPa (12 kbar) and several hundred degrees, a series of black phosphorus modifications are formed that are characterized by even higher densities (2.70 g cm³). These include orthorhombic, rhombohedral, and cubic varieties. The black forms have lower reactivity and solubility than red phosphorus.

Phosphorus vapor exists as P₄ molecules until dissociation to P₂ begins at 800°C. Essentially all the vapor is P₂ at 1500°C, but further dissociation to monatomic P is less than 0.1% at that temperature.

Chemical Properties

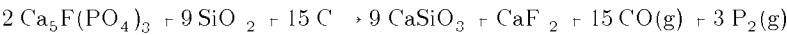
Phosphorus shows a range of oxidation states from -3 to +5 by virtue of its electronic configuration. Elemental P₄ is oxidized easily by nonmetals such as oxygen, sulfur, and halides to form compounds such as P₂O₅, P₂S₅, and PCl₃. It is also reduced upon reaction with metals to generate phosphides. The phosphides may be quite inert when formed with transition metals, for example, iron(III) phosphide, FeP, or may be chemically reactive when formed with active metals like aluminum or sodium, AlP or Na₃P. In some reactions phosphorus may be both oxidized and reduced, as in reactions with metal salt solutions like copper sulfate, where both phosphoric acid and copper phosphide are generated. Certainly the more useful and common reactions of phosphorus involve oxidation with air, sulfur, or chlorine. The largest volume products are phosphoric acid and phosphate derivatives of phosphoric acid (see PHOSPHORUS COMPOUNDS).

Elemental phosphorus reacts with oxidizing acids such as nitric or strong sulfuric. It also reacts with alkali, forming a combination of phosphine, hypophosphite, and phosphite at increasing rates as the pH increases. In reactions related to electric furnace operation, P₄ is not reactive with carbon monoxide over a wide range of temperatures as shown by its compatibility in furnace gases. However, at high temperatures over 1000°C, P₄ reacts with steam to generate P₂O₅ and hydrogen (8).

Manufacture of White Phosphorus

As of the mid-1990s all commercial phosphorus is manufactured at a few sites around the world. Significant production occurs in Idaho and Montana in the United States, in the Netherlands, in Kazakstan, and in China; smaller production occurs in France, Russia, and India. A large amount of furnace capacity has been shut down worldwide because of cost pressure from electric power costs, phosphate derived from purified wet acid, and detergent phosphate bans legislated in the Western World. However, as of late 1995, additional production is still being brought on line in China.

Elemental phosphorus is produced from a phosphorus-rich ore mostly recovered by strip mining. This ore usually contains fluorapatite, plus some silica and silicates. When a carbon source, usually coke, is added to the ore at temperatures greater than 1100°C, the following overall reaction occurs:



At the lowest reaction temperatures, P₄ tetrahedra rather than P₂ dimers may be produced, or as the gas cools, dimers combine as:



There are several theories about the course of this reaction, but the actual reaction mechanism in the electric furnace is not proven and may vary depending on the location in the furnace. The reaction is endothermic and does not occur at significant rates except at elevated temperatures. The reaction is facilitated by adding additional silica beyond that contained in the ore, usually as the mineral quartzite. The silica serves a twofold purpose. First, it lowers the slag melting point, and second, it combines with the lime values in the ore to improve the thermodynamics of the reaction. Controlling the melting point of the slag is critical for optimizing phosphorus production. If reaction temperatures are increased over 1700°C, unwanted side reactions such as the reduction of silica become excessive. These reactions unproductively consume energy and can contaminate the phosphorus product. The most productive reaction temperature range appears to be 1400–1600°C.

The most effective phosphorus production technology uses a submerged arc furnace. The submerged arc furnace performs three functions: chemical reactor, heat-exchanger, and gas–solid filter, respectively, each of which requires a significant amount of preparation for the solid furnace feed materials.

The two ore preparation technologies in common use are moving grate calciners and rotary kilns. The primary purposes of this preparation step are to produce strong feed agglomerates, often called nodules; to provide a consistently sized material; and to remove energy-consuming impurities that disrupt furnace operation. Some common preparation steps for the silica include crushing, screening, washing, and drying. The coke may be metallurgical, petroleum, or formed coke, but must be size-controlled and is often dried before use.

Once each of these materials has been prepared for furnace use, a furnace charge (burden) is produced by mixing and proportioning the three components. The mixing process is accomplished using accurate weighing systems. The burden is then transported to the furnace charge bins. As the furnace is operated, feed falls continuously by gravity from the charge bins through a system of vertical chutes into the furnace. Two molten furnace by-products, slag and ferrophos, are tapped (removed) from the furnace using either interval or continuous tapping. The furnace off-gas containing primarily phosphorus and carbon monoxide leaves the furnace by flowing up through the porous burden while exchanging heat and dust. Next, the off-gas is usually cleaned of dust particles, using an electrostatic precipitator. Lastly, P₄ is separated from the CO using a water-spray condenser, and the molten P₄ is pumped to storage tanks. Each of these process steps is discussed in greater detail in the following sections. The general steps in the process are shown schematically in Figure 1.

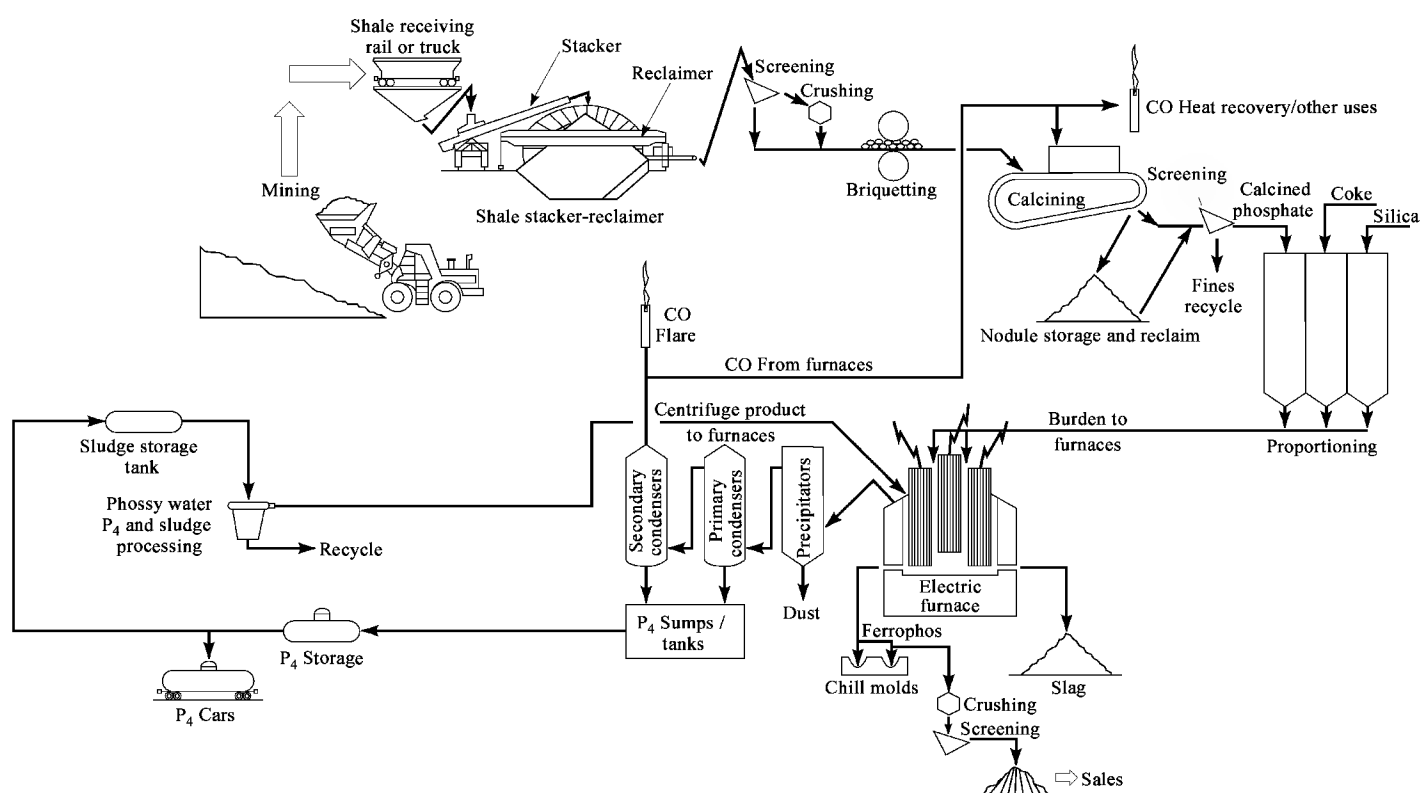


Fig. 1. Schematic of the elemental phosphorus process.

Mining. With the closure of all North American white phosphorus production facilities outside of the western United States, the only remaining mines utilized for U.S. P_4 production are located along the southern Idaho–Wyoming border. The three remaining white phosphorus producers as of mid-1995, Monsanto (Soda Springs, Idaho), Rhône-Poulenc (Silver Bow, Montana), and FMC (Pocatello, Idaho), all operate mines near Soda Springs, Idaho. However, some ore mined in Florida is still shipped overseas for use in elemental phosphorus plants.

The western phosphates are sedimentary deposits in adjoining areas of Wyoming, Idaho, and Utah derived from a former inland sea. They consist of layers of limestone, phosphate, and chert, now buckled and faulted so they are rarely horizontal. The phosphate ore is strip-mined using large earth-moving equipment such as shovels, scrapers, dump trucks, and bulldozers to mine the overburden and phosphate ore. Mining ratios of overburden to metric ton of recovered ore are from 1–3 m³ t (2–4 yd³ short ton). The typical mining practice is to remove ore and overburden from a pit in discrete layers (lifts) of 10–20 m in depth. Overburden from the pit is back-hauled to a previously mined pit. Extensive land reclamation practices are later carried out to return the mine areas to natural states.

The ore is stockpiled at the mine in several piles based on the particular operating strategy of assay and ore chemistry. Typical assay of the blended grades used in the process is about 23–30% P_2O_5 , 35–40% CaO , and 20–30% SiO_2 . Owing to the cold winters of Montana and eastern Idaho, the ore is usually shipped to the plants only during the warm part of the year, between April and October. This eliminates the need for ore car heating to remove frozen ore from the cars. Transportation methods include both open railcars and large trucks, depending on the mine and plant locations and associated economics.

Feed Preparation. The burden charged to a phosphorus furnace must be kept porous enough to allow the gases generated in the reaction zone near the bottom of the furnace to escape while losing heat to the feed and being cleaned of entrained dust. To allow sufficient porosity, the diameter of the phosphate ore, quartzite, and coke particles are sized in the range of 0.5–5 cm and must not contain excessive fines that will block the gas flow. The coke is received from suppliers already in the appropriate sizes or is crushed and screened on-site. The quartzite is generally mined locally and is crushed and screened to obtain the desired size distribution. The phosphate ore usually consists of fine particles that must be agglomerated and sintered by some process into a hardened mass that will resist deterioration during subsequent handling. The sintering or calcination step also serves to remove organics and other impurities that contaminate the product and cause bridging of the burden above the reaction zone in the furnace, which again restricts the gas flow and inhibits the heat exchange and dust removal from the off-gas. A number of technologies are therefore utilized in the phosphorus industry to prepare the phosphate ore.

One ore preparation method employs a grate calciner similar to a Dwight-Lloyd Sintering Machine. This machine consists of a continuous belt of connected cast iron grate modules that move through various heating and cooling zones on tracks. In this process the fine phosphate ore is first pressed into green briquettes similar in appearance to charcoal briquettes (Fig. 2). The briquettes are fed onto the grate to form a level bed of material. Combustion of carbon monoxide from the furnace off-gas, supplemented by other fuel, is used to heat an air stream that is pulled down through the briquettes. The temperature of the bed can reach to above 1200°C. This process forms a durable ore nodule which can undergo additional handling without significant deterioration prior to being charged into the furnace.

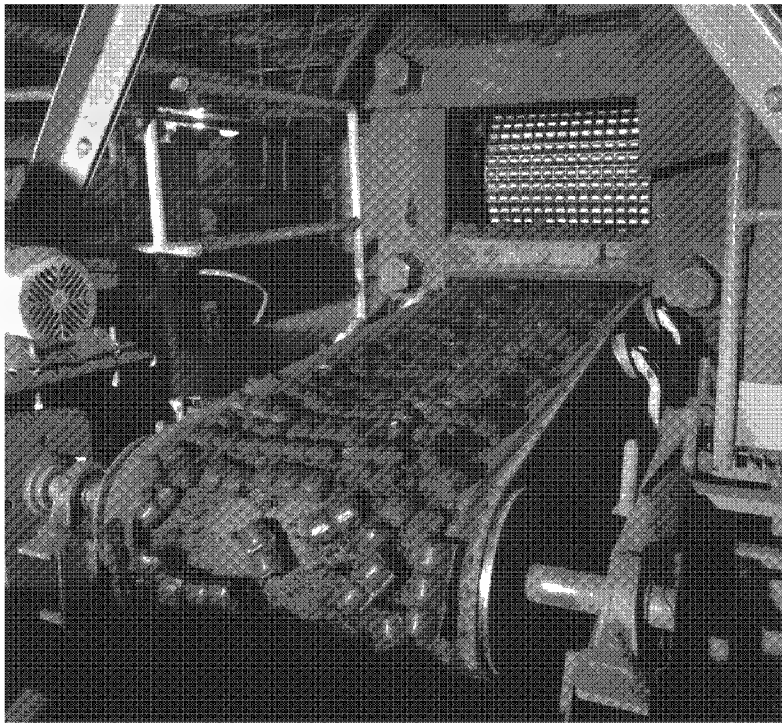


Fig. 2. Phosphate ore briquetting.

Another preparation method is a sintering process where phosphate ore, sand, and coal are blended together and ignited on the grates of a sintering machine. Air is pulled through the blend, and the entire mass is allowed to burn. The resulting fused bed of material is then crushed and screened to the appropriate size distribution, and the undersized material is reprocessed.

Still another ore preparation is the nodulizing process where the ore is heated in a rotary kiln to incipient fusion. The tumbling action in the kiln causes the phosphate ore to cohere and form spheroidal agglomerates. Combustion of carbon monoxide from the furnaces is used along with supplemental fuel to supply heat to 1300–1500°C. A boring bar is used near the kiln discharge to aid in breaking up the fused ore. The material is then cooled, crushed, and screened to the appropriate size for furnace feed.

One other method used to size the phosphate ore is disk agglomeration. After preparation, the disk agglomerates are sintered at high temperature in separate process steps, followed by screening to the appropriate size specifications.

Some producers beneficiate the phosphate ore prior to agglomeration to increase the phosphate content and remove undesirable contaminants. One approach uses a water wash to remove fines which are lower in assay and higher in contamination. If too much clay is removed from the ore by the beneficiation, a binder such as bentonite clay must be added back to the ore to facilitate agglomeration.

Electric Furnace. Two basic types of phosphorus furnaces based on electrode configuration are in use. First, the in-line type having three electrodes lined up in a row allows for a simpler layout and a rectangular furnace design at the expense of electrical energy efficiency. Single-phase transformers incorporating Y or delta electrical connections having power ratings of 15–40 MW are typical of operating in-line furnaces. Second, the delta type of furnace uses a three-electrode, symmetrical triangular configuration with a three-phase Y or delta electrical connection operating in the 45–65 MW range at potentials of 200–650 V. World class phosphorus furnaces are of this type. The delta furnace is circular, polygonal, or of a rounded triangular cross-section type. A typical delta furnace along with accompanying off-gas handling system is shown in Figure 3. The furnaces are basically run continuously, except during repairs, process upsets, power curtailments, or electrode building. Typical operating characteristics for a large (60-MW), phosphorus furnace are shown in Table 2.

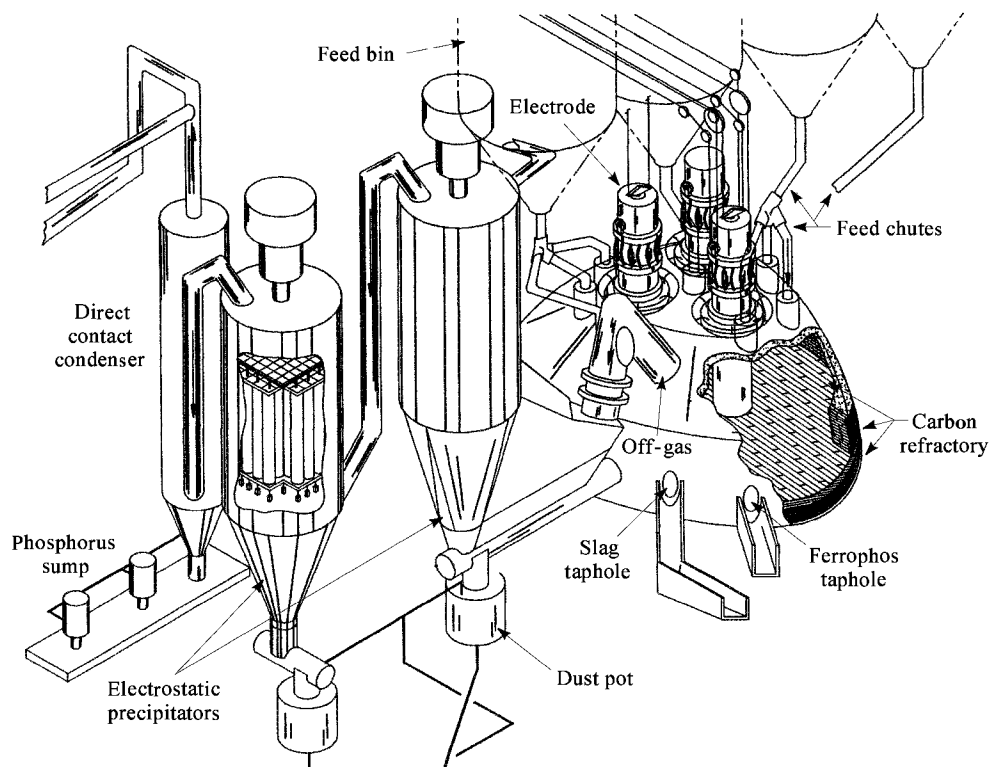


Fig. 3. Elemental phosphorus furnace.

Table 2. Operating Characteristics of a 60-MW Phosphorus Furnace

Parameter	Value
average potential between electrodes, V	250–350
power factor	0.96–0.98
raw materials consumed ^a	
power, kWh	13–15
fluorapatite ore, kg	10–13
coke, kg	1.2–1.7
silica, kg	1–2
products formed ^a	
slag	8–9
ferrophos	0.1–0.2
furnace off-gas	2.6–2.9
recovery, based on P ₄ charged to the furnace, %	84–90
temperature of off-gas, °C	300–450
temperature of slag at tapping, °C	1400–1500

^a Per kilogram of elemental P₄ produced.

The electrodes are cylindrical and hang vertically in the furnace from suspension cables or hydraulic, water-cooled electrode clamps, which allow up-and-down movement of the electrode via an automated amperage control system. Prebaked carbon–graphite electrodes are used exclusively in domestic furnaces, with some foreign facilities utilizing Sinderberg electrodes. Electrodes of 110–140 cm diameter and a nominal 2.5-m length are generally used. A tapered male–female thread design is used to build electrodes on top of one another as they are consumed in the furnace at about 15–30 kg per 1000 kg phosphorus. Substantial mechanical and thermal stresses are imposed on the electrodes during operation and can result in electrode breakage which can severely affect furnace operation and on-stream time. Proper attention to assuring electrode quality, joint integrity, and building procedures is essential. Electrodes carry up to 65 kA, which necessitates careful attention to electrical design in order to optimize the furnace power factor, which is about 0.96–0.98. Some producers have developed hollow electrode technology to convey carbon–coke fines directly into the reaction zone at the tip of the electrode, which results in decreased electrode consumption.

The outer shell of the furnace (roof, sides, and bottom) is of welded steel plate construction employing various alloys to meet the electrical, thermal, and corrosion requirements at specific locations. Water and air cooling are used extensively to control refractory and metal temperatures. Large carbon blocks cemented together with carbon paste or ramming material are used to construct a monolithic bottom hearth 1–1.5 m thick. Carbon and graphite bricks, rams, and pastes are used as the primary refractory materials in the sidewall to just above the molten slag zone of the furnace. Fire brick is typically used on the upper sidewalls with castable and gunnite high alumina refractory used on the furnace roof. The electrodes enter the furnace through special refractory lined sleeves that are sometimes water-cooled. A telescopic seal is customarily used to provide a gas seal between the electrode and the furnace, which operates at about ±23 cm water pressure. Refractory performance is critical in obtaining satisfactory furnace operability, and is key to getting long life between furnace rebuilds, which is a significant cost factor of P₄ manufacturing.

The raw materials, or burden, are fed by gravity from surge bins above the furnace via choke-fed feed chutes that distribute burden evenly around each of the three electrodes. It is important in the design of the feed delivery system to minimize material segregation, feed pluggages, and size degradation in order to maintain optimal furnace operation. It is common practice to inject inert gas into the feed chutes to minimize the chance of air in-leakage into the furnace or furnace off-gas leakage to the feed bins. Hollow electrode technology can also result in improved raw material utilization and furnace control. Control of the burden sizing and chemistry is necessary to control the reaction zone position and burden porosity, which is a key factor in optimizing P₄ recovery efficiency and off-gas temperature, off-gas flow, and dust carryover.

The phosphate reduction occurs in the high (1300–1600°C) temperature reaction zone in the lower portion of the furnace. Slag and ferrophos are two molten by-products produced from the reduction reaction. Ferrophos, which is more dense than slag, is tapped from one or more tapholes about 5–8 cm in diameter near the top of the carbon hearth refractory. The slag is tapped from one or more tapholes about 5–8 cm in diameter nominally 25 cm above the ferrophos taphole. Both tapholes are plugged with clay or other suitable material. The slag taphole is generally water-cooled to extend the life of the surrounding refractory. Slag is tapped almost continuously from the larger furnaces. Three principal slag-handling techniques are (1) tapping into pits which are cooled via water sprays, followed by removal with front-end loaders; (2) tapping into ladles which allow transport of the molten slag to remote slag dumps; and (3) granulating with high volumes of water to rapidly cool the slag into small-sized aggregate. Ferrophos is tapped 1–3 times per day, depending on iron concentrations in the burden. Care must be taken to avoid and protect against contact between molten ferrophos and water because of the likelihood of steam explosions, which can be extremely hazardous. Ferrophos tapping is into pots or molds where it is allowed to cool and is later crushed and sized for sale for metal recovery processing or alloy additives.

Product Recovery. At standard conditions (25°C, 101.3 kPa (1 atm)) typical off-gas compositions are about 86% CO, 7.5% P₄, 5% H₂, 1% N₂, and traces of PH₃, CO₂, F, and S; large furnaces generate off-gas at a rate of about 120–180 m³ min. In most installations the off-gas is passed through a series of Cottrell electrostatic precipitators which remove 80–95% of the dust particles. The precipitators are operated at temperatures above the 180°C dew point of the phosphorus. The collected dust is either handled as a water slurry or treated dry. Final disposal is to a landfill or the dust is partially recycled back to the process. The phosphorus is typically condensed in closed spray towers that maintain spray water temperatures between 20 and 60°C. The condensed product along with the accompanying spray water is processed in sumps where the water is separated and recycled to the spray condenser, and the phosphorus and impurities are settled for subsequent purification.

Although most of the particulate in the off-gas from the furnace can be captured by the electrostatic precipitators before condensing the phosphorus, some carryover into the product P₄ is inevitable. This particulate is partly separated into the condenser water. The remainder reports to the phosphorus to yield either dirty product or a stable emulsion called phosphorus mud or sludge. Over many years a variety of approaches have been used to minimize the formation of sludge and to recover phosphorus product from the sludge.

Less stable parts of the sludge can be treated by holding in tanks for extended periods of time to allow the weaker emulsion to break and separate a clean P₄ product. The more stable sludges can be broken by mechanical action in filters or centrifuges, by recycle to the furnace for redistillation, or by redistillation in auxiliary units. Chemical attack via oxidation or complexing agents that break the emulsion has also been employed.

Product Quality and Specifications. Most of the elemental phosphorus produced is converted to derivatives by the manufacturer. Some white phosphorus is sold on the open market. Typical manufacturers' phosphorus analyses for a straw yellow product follow:

phosphorus, wt %	>99.90
arsenic, ppm	30–250
toluene insolubles, wt %	0.01–0.05
oils (hydrocarbons), ppm as carbon	30–100

Trace metals are analyzed and kept at levels appropriate to the special needs of the customer.

For analysis, white phosphorus is typically extracted through a fritted thimble with refluxed toluene. Any trace amounts of water are captured in a calibrated sidearm to the apparatus. The solids on the frit are weighed, the water measured, and the phosphorus calculated by difference. For impure samples of phosphorus, the toluene extract may be analyzed with a gas chromatograph (gc) equipped with a phosphorus–nitrogen detector.

Trace contaminants in the phosphorus may be determined by oxidation of the phosphorus by various techniques. The metals are then determined by an inductively coupled plasma spectrophotometer or by atomic absorption. The most important trace metal is arsenic, which must be reduced in concentration for food-grade products. Numerous other trace metals have become important in recent years owing to the specifications for electronic-grade phosphoric acid required by the semiconductor industry (see ELECTRONIC MATERIALS; SEMICONDUCTORS). Some trace elements must be reduced to the low ppb range in phosphoric acid to comply.

Analyses of raw materials and by-products from the phosphorus process are generally done by x-ray florescence or other instrumental methods. Statistical process control is utilized to ensure that critical process parameters are maintained in control. Key analytical parameters that are monitored include P₂O₅ and ignition loss in the ore; Fe₂O₃, SiO₂, CaO, and sizing in all raw materials; fixed carbon in coke; and P₂O₅ and SiO₂ CaO ratio in the slag, to name a few. Very few classical wet methods are employed except for occasional nonroutine samples or for instrument standardization.

Shipping and Handling

Phosphorus is stored and handled under a protective layer of water. Production quantities are transferred as a liquid by either water displacement or pumps, with water recycle to maintain the water balance and cover. In earlier times, phosphorus was sometimes stored in underground tanks or pits, but as of the 1990s storage is limited to tanks located inside diked areas that are accessible on the outside for safety and leakage control.

For off-site transportation, the phosphorus is loaded into railcars for transfer to the sites where it is used directly as a raw material or burned and hydrated to phosphoric acid. During shipping, the phosphorus is allowed to solidify in the cars. The railcars are commonly double walled with a jacket that can be heated with steam or hot water so that the phosphorus can be remelted on-site for transloading to local storage tanks. For overseas shipping, tanktainers with reinforced superstructure for safe handling are used. Formerly, full tanker ships were in use.

Smaller amounts of phosphorus, or elemental phosphorus-containing materials, are also shipped in 115-L (30-gal) drums that are DOT regulated (U.S. DOT 1A1 or 1A2 classification) and have thick shells and special gaskets and fittings for protection. Quantities up to 0.5 kg (1 lb) are allowed for shipping in two hermetically sealed (soldered), nested cans inside a wooden box where the empty space is filled with vermiculite (U.S. DOT 4C1, 4C2, 4D, or 4F classification). All air transportation of elemental P₄, both U.S. and international, was prohibited beginning in 1992.

A DOT regulation covers both domestic and international shipping (9). For transportation safety, the DOT has information for first responders to incidents involving elemental phosphorus (10). In addition, the Chemtrec phone number 1-800-424-9300 accesses DOT emergency information and assistance in the United States. Also, the phosphorus producers in the United States have established a Phosphorus Emergency Response Team (PERT) to assist in handling P₄ emergencies.

Health and Safety

At ambient temperatures white phosphorus spontaneously ignites when exposed to air. It has an autoignition temperature of 30°C. As a result, any human exposure to white phosphorus can cause severe thermal burns to the skin and eyes. The vapor from phosphorus can cause severe lung irritation, followed by a build-up of fluid in the lungs. Continuous long-term inhalation of white phosphorus vapor (>0.1 mg m⁻³) can result in bone loss to the jawbone structure causing loosening of teeth and severe pain and swelling of the jaw (11). This condition is commonly referred to as phossy jaw. Some evidence exists that increased infant mortality can result when pregnant women are exposed to P₄ vapor in excess of 0.075 mg kg⁻¹ day (12). Ingestion of white phosphorus is potentially fatal. The lowest reported fatal dose is 1 mg kg⁻¹ for humans (11). Absorption through the skin is also possible but is only considered to be a moderate hazard compared to the other routes of exposure.

When exposed to air, white phosphorus oxidizes to phosphorus pentoxide forming copious quantities of white smoke. This smoke may be irritating but is not considered to be toxic. White phosphorus solid, liquid, or vapor is also extremely reactive with oxidizers such as strong acids, alkaline hydroxides, halogens, and nitrates. Contact of phosphorus with water or oxidizers also generates phosphine [7803-51-2], PH₃, a highly toxic and flammable gas. Phosphine has an 8-h time-weighted average exposure limit of 0.3 ppm (13). Under alkaline conditions the rate of PH₃ formation is high. At neutral or acidic pH, the PH₃ generation is slow but still very hazardous if the PH₃ is allowed to accumulate in a confined vapor space. The safest commercial handling conditions for molten phosphorus are generally considered to be from pH 6 to 8 at 45–65°C.

Phosphorus production plants and users should ensure that processing of the material is contained and that potential high exposure areas are well ventilated. Workers must wear aluminized fiber glass or Kevlar flame-retardant full protective clothing, face shield with hard hat, rubber boots, and heavy rubber gloves when handling or transferring the product. The phosphorus should always be kept under neutral pH water at temperatures less than 65°C or under an inert atmosphere to avoid oxidation and exposure hazards. High potential exposure areas should also be equipped with well-maintained, water-filled safety tubs, deluge systems, and water-spray extinguishing systems as a precautionary measure. If high exposure levels to phosphorus vapors or phosphine are anticipated, then self-contained breathing apparatus units should be utilized. Individuals exposed to phosphorus through skin or eye contact should have the exposed area flushed immediately with large amounts of water. The affected area should be kept wet until all of the phosphorus is removed or flushed away. Victims of phosphorus inhalation should immediately be removed to an area with fresh air and have artificial respiration administered, if necessary. Workers who have had dental surgery and pregnant women should be kept away from phosphorus exposure areas completely. Anyone who ingests phosphorus should drink a large volume of water and be induced to vomit. Medical assistance should be obtained as soon as possible after any instance of phosphorus exposure.

By-Products

The electric furnace process generates four streams that can be considered by-products: slag, ferrophos, precipitator dust, and carbon monoxide off-gas. The approximate composition of the slag and precipitator dust are given in Table 3. These vary somewhat among different phosphorus manufacturers.

Table 3. Composition of Phosphorus Manufacture By-Products, Wt %

By-product	P ₂ O ₅	CaO	SiO ₂	Al ₂ O ₃	F	Fe ₂ O ₃	K ₂ O	ZnO	C
slag	1.0–2.5	40–50	38–44	3–7	2–3	0.1–0.5			
precipitator dust	20–30	5–15	15–30	2–4	3–5		5–20	5–15	0–10

The slag has the primary chemical composition of pseudowollastonite, CaSiO₃, having significant contents of calcium fluoride and alumina, and other components depending on the ore source. This slag has been used as railroad ballast and road and building-block aggregate. Application is clouded as of this writing by questions about the low level radioactivity common in western U.S. ores and slags. Most furnace slag has been accumulated at plant sites, where it is considered inert and nonhazardous.

Ferrophos consists primarily of 50–60 wt % iron and 18–28 wt % phosphorus in proportions representing Fe₂P plus smaller amounts of either FeP or Fe₃P. However, varying amounts of chromium (4–5 wt %) and vanadium (5–6 wt %) present in the ore are also included in the ferrophos, and are recovered in some cases. From 0–2 wt % of silicon may also be present. Ferrophos is also sold for use in specialty steels that can accommodate the phosphorus content in their formulation (see STEEL).

Precipitator dust often contains concentrated amounts of minor ore components that make it attractive. The potassium, phosphate, and zinc content have resulted in its use in fertilizer, and the silver and gallium content have been the subject of some recovery efforts (see RECYCLING).

The carbon monoxide off-gas which is saturated with H₂O and P₄ after condensation of the phosphorus is usually used as a fuel for calcination of the ore in the front end of the process. This gas contains 90–93% CO, 1–3% N₂, 1–8% H₂, and 0–1% CO₂.

Environmental Control

Pollution control has become a primary concern and expense of elemental phosphorus producers since the 1970s. Problem areas include mining dusts, feed preparation and feed-handling dusts, slag dusts and emissions, precipitator dust, off-gas flare emissions, and phossy water and sludge treatment. The specific contaminants include elemental phosphorus, dust particulates, leachable hazardous metals, and low level radioactivity. These problems are being addressed by a variety of technical options which are often unique to each process site. The final discharges from the phosphorus plants are being confined to slag, ferrophos, and precipitator dusts, which are all nonhazardous wastes. Commercial application of the slag and precipitator solids is a point of contention owing to low level radioactivity. Ferrophos is often sold for its iron, vanadium, and chromium content.

Economic Aspects

Beginning in 1969, a movement to restrict and then legally ban the use of phosphates in detergents led to the closing of significant amounts of plant capacity (Table 4). The movement derived from an unproven belief that eutrophication of lakes and streams could be reduced by eliminating the phosphates contained in home laundry detergents. Phosphorus producers objected that only 3% of the phosphorus in U.S. rivers, lakes, and streams was contributed by detergent phosphates. The remaining 97% of the phosphorus comes mostly from agricultural fertilizer run-off, plus human and animal waste (15). Nevertheless, legal bans on phosphate detergents were instituted in enough locations to effectively eliminate phosphate detergents in most of the United States. Similar legislative bans have been enacted in some European countries.

Table 4. North American Electric Furnace Phosphorus Capacity, t × 10³/yr^a

Producer	1964	1970	1975	1981	1988	1994
FMC	68	132	132	132	122	122
Monsanto	136	222	222	209	100	104
Rhône-Poulenc (Stauffer)	111	116	102	95	68	38
Occidental (Hooker)	68	63	52	52	52	
Mobil	67	53	18	18		
TVA	33	36	33			
Total U.S.	483	622	559	506	342	264
Albright & Wilson (Canada)	25	71	71	89	89	
Total, North America	508	693	630	595	431	264

^a Refs. 6, 14.

A second pressure on elemental P₄ production was the development of processes which remove impurities from phosphoric acid made by the wet process, to generate acid of equivalent purity to that obtained by the electric furnace route. Two such plants were brought on stream: one at Aurora, North Carolina in 1990 by a joint venture of Albright & Wilson, Texasgulf, and Olin, and another at Geismar, Louisiana, in 1991 by Rhône-Poulenc. These units have reported capacities of 47,600 and 31,700 t yr elemental phosphorus equivalent, respectively (14).

Estimated world production capacity for elemental phosphorus is shown in Table 5 (6). Three elemental phosphorus production sites remain operational in North America (14), although Rhône Poulenc has announced its intention to cease production in late 1995. The remaining plants have survived owing to the availability of economical electric power in the Northwest and proximity to phosphate ore deposits, resulting in lower cost phosphorus. The capacity of these producers in 1995 was estimated to be 264,000 metric tons. U.S. production capacity peaked at approximately 622,000 metric tons in 1970.

Table 5. Estimated 1995 World Capacity of Elemental Phosphorus^a

Region	Company	Plant location	Approximate annual capacity, ^b t × 10 ³ yr
United States	FMC	Pocatello, Idaho	122
	Monsanto	Soda Springs, Idaho	104
	Rhône-Poulenc	Silver Bow, Montana	38
	Total, U.S.		264
Europe	Atochem SA	Epierre, France	15
	Hoechst	Huerth-Knapsack, Germany	(30)
		Vlissingen Holland	91
	VEB Stickstoffwerk	Piesteritz, Germany	(20)
	Enimont Augusta	Crotone, Italy	(15)
	Industriale Sd		
	Total, Europe		171
Eurasia	RAO Phosphoz	Togliatti, Russia	36
	Nodfos	Dzhambul, Kazakstan	(115)
	Nodfos	Nova Dzhambul, Kazakstan	181
	Phosphor	Chimkent, Kazakstan	172
	Liuchen	Guang, Xi, China	15
	unknown	Yunan Province, China	109
	unknown	other locations, China	172

	Excel Industries	Bhavnagar, Gujarat, India	5
	Star Chemicals	Thana, Maharashtra, India	3
	United Phos	India	4
<i>Total, Eurasia</i>			<i>812</i>
<i>Total world capacity</i>			<i>1247</i>
<i>Total world active capacity</i>			<i>1067</i>

^a Ref. 6.
^b Numbers in parentheses are those of idled plants.

Outside of North America, the total number of plants, not including an undetermined number of smaller plants in China, has declined to about a dozen. Four are in Russia and Kazakstan, where the future viability of the facilities is uncertain. Three of these four plants were operating at considerably below their 505,000 metric ton capacity in 1995. The plant at Dzhambul, Kazakstan is believed not to be in operation at this writing. Meanwhile, China, reported to have considerable total capacity in their smaller plants, is building two large furnaces using Russian technology. The five phosphorus plants that remain in Western Europe have a production capacity of 171,000 metric tons, but three are believed idled. This is down from 205,000 metric tons produced in 1980 and can be attributed, in part, to the phosphate bans enacted during the past decade in several Western European countries.

The price of white phosphorus has been as low as \$1.59 /kg and as high as \$4.91 /kg since the mid-1980s. However, through March 1995 the list price had remained stable for more than a year at \$2.00 /kg, freight equalized, in tank cars. After rationalization stabilizes the industry, the expectation is that the demand for white phosphorus should grow 2% annually through 1998 (14).

The cost of producing elemental white phosphorus can be broken down into four categories: power, 25.0%; raw materials, which include ore, coke, and silica, 30.0%; labor, 20.0%; and others, which include capital costs, electrodes, fuel, supplies, etc, 25.0% (16).

Uses

About 264,000 metric tons of elemental P₄ capacity is available in North America, plus another 79,000 t (P₄ equivalent) of purified wet phosphoric acid (14). About 85% of the elemental P₄ is burned to P₂O₅ and hydrated to phosphoric acid. Part of the acid (ca 21%) is used directly, but the biggest part is converted to phosphate compounds. Sodium phosphates account for 47%; calcium, potassium, and ammonium phosphates account for 17%. Final applications include home laundry and automatic dishwasher detergents, industrial and institutional cleaners, food and beverages, metal cleaning and treatment, potable water and wastewater treatment, antifreeze, and electronics. The purified wet acid serves the same markets.

The remaining 15% of the elemental P₄ is used in P₄-dependent applications which require the element as a direct reactant. The principal products include P₂S₅, PCl₃, and POCl₃, P₂O₅, and hypophosphite, with much smaller amounts leading to PH₃, red P, phosphonates, and various other phosphorus derivatives. Final applications include flame retardants (qv), lubricant additives, insecticides, herbicides, water treatment, cleaning compounds, plasticizers, and semiconductors (14).

Manufacture of Red Phosphorus

Red phosphorus is manufactured from white phosphorus for applications such as striking surfaces for matches, fireworks (see PYROTECHNICS), flame retardants in polymers, semiconductors, and PH₃ used to manufacture semiconductors. Manufacturers include Hoechst (Germany), United Phosphorus (India), Nippon (Japan), Rinkagaku Kogyo (Japan), and Italmatch Srl (Italy). The batch conversion process requires one to two days and involves heating molten white phosphorus at temperatures somewhat below its boiling point of 280°C. The conversion is exothermic, so the reaction temperature must be controlled to prevent accidents. About half of the added white phosphorus is converted before the accumulation of solid red P in the molten phosphorus drives the viscosity of the melt to a thick slurry. Solidification occurs at about 70% conversion to red phosphorus. After most of the conversion takes place, the reaction may be driven to completion at higher temperatures of 300–600°C, where the remaining white P₄ is converted or vaporized. Alternatively, the remaining P₄ can be reacted with alkali. If the particle size is not controlled during reaction, the final product is then wet-ground or ball-milled and finally dried.

Continuous processes and routes involving added pressure and shorter conversion times have been developed. The process may also be adapted either to add stabilizers to reduce the final reactivity of the red P, or to ensure high reactivity such as for subsequent conversion to PH₃.

BIBLIOGRAPHY

"Phosphorus and Phosphorus Alloys" in *ECT* 1st ed., Vol. 10, pp. 442–461, by J. R. Van Wazer, Monsanto Chemical Co.; "Phosphorus and the Phosphides" in *ECT* 2nd ed., Vol. 15, pp. 276–295, by J. R. Van Wazer, Monsanto Co.; in *ECT* 3rd ed., Vol. 17, pp. 473–490, by J. R. Van Wazer, Vanderbilt University.

1. J. R. Van Wazer, *Phosphorus and Its Compounds*, Vol. II, Wiley-Interscience, New York, 1961.
2. H. Diskowski and T. Hofmann, in *Ullmann's Encyclopedia of Industrial Chemistry*, 5th ed., Vol. A 19, VCH Verlagsgesellschaft, Weinheim, Germany, 1991, pp. 505–525.
3. D. C. DeWitt, in *Encyclopedia of Chemical Processing and Design*, Vol. 36, Marcel Dekker, Inc., New York, 1991, pp. 1–33.
4. R. B. Burt and J. C. Barber, *Production of Elemental Phosphorus by the Electric-Furnace Method*, Chemical Engineering Report No. 3, Tennessee Valley Authority, Wilson Dam, Ala., 1952.
5. J. R. Van Wazer, *Phosphorus and Its Compounds*, Vol. I, Interscience Publishers, Inc., New York, 1958.
6. "Phosphorus and Phosphorus Chemicals Report," *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., 1991, 1995, pp. 760.2000A–760.2001R.
7. A. E. G. Robiette, *Electric Smelting Processes*, Charles Griffin and Co., Ltd., London, 1973, pp. 254–271.
8. J. F. Shultz and co-workers, *Ind. & Eng. Chem.* **42**, 1608 (1950).
9. "Phosphorus, White, Under Water," *Code of Federal Regulations*, Title 49, U.S. Department of Transportation, Washington, D.C., Oct. 1994.
10. *Emergency Response Guide* (ERG), Guide No. 38, U.S. Department of Transportation, Washington, D.C., 1993.
11. *American Conference of Governmental Industrial Hygienists (ACGIH)*, Vol. II, Cincinnati, Ohio, 1991, p. 1252.
12. E. C. Robinson and co-workers, *The Toxicologist* **11**, 112 (1991).
13. Ref. 11, p. 1248.
14. "Chemical Profiles: Phosphorus," *Chemical Marketing Reporter*, Schnell Publishing Co., Inc, New York, 1964–1995.
15. *Chem. Mktg. Rep.*, 23 (May 29, 1989).
16. Technical data, FMC Corp., Pocatello, Idaho, 1994.

J. R. Brummer
J. A. Keely

T. F. Munday
FMC Corporation

PHOSPHORUS COMPOUNDS

Phosphorus compounds exhibit an enormous variety of chemical and physical properties as a result of the wide range in the oxidation states and coordination numbers for the phosphorus atom. The most commonly encountered phosphorus compounds are the oxide, halide, sulfide, hydride, nitrogen, metal, and organic derivatives, all of which are of industrial importance. The halide, hydride, and metal derivatives, and to a lesser extent the oxides and sulfides, are reactive intermediates for forming phosphorus bonds with other elements. Phosphorus-containing compounds represented about 6–7% of the compound listings in *Chemical Abstracts* as of 1993 (1).

The largest-volume phosphorus compounds are the phosphoric acids and phosphates (qv), ie, the oxide derivatives of phosphorus in the +5 oxidation state. With the exception of the phosphoric acid anhydride, P₄O₁₀, and the phosphate esters, these materials are discussed elsewhere (see PHOSPHORIC ACIDS AND PHOSPHATES). An overview of phosphorus compounds other than the phosphoric acids and phosphates is given herein. These compounds constitute a large variety of phosphorus compounds that are either nonoxide derivatives or derivatives of phosphorus in oxidation states lower than +5. These phosphorus compounds are manufactured only from elemental phosphorus (qv) obtained by reduction of naturally occurring phosphate rock (calcium phosphate).

Because of the high stability of the P–O and P=O bonds, the largest group of phosphorus compounds in existence is the oxides. The oxyacids form the basis for the most systematic nomenclature (2,3). Table 1 lists the well-characterized lower molecular weight oxyacids and the corresponding structures. The basicity of the acid is related to the P–OH moiety providing the acid function. Whereas phosphonic acid is often referred to as phosphorous acid, the free acid of formula H₃PO₃ exists almost exclusively in the form of the stable, well-characterized phosphonic acid, HP(=O)(OH)₂, and not as the structural isomer, phosphorous acid [10294-56-1], P(OH)₃. Phosphinic acid, H₃PO₂, is often called hypophosphorous acid. Some structures may be found primarily as salts or esters. The acid forms are encountered infrequently if at all. Examples include the esters of phosphinous acid [25756-87-0], H₂POH, and the phosphite esters, P(OR)₃, the latter structurally derived from phosphorous acid. Acids and salts containing more than one phosphorus atom of the same or different oxidation state are also known, such as diphosphoric(III,V) acid [14902-77-3], H₄P₂O₇. These lower oxidation state acids or salts containing more than one phosphorus atom are encountered infrequently and are often formed, eg, as somewhat metastable intermediates in the hydrolysis of phosphorus halides. They may possess either P–P or P–O–P linkages. Table 2 provides information for organo derivatives of the phosphorus oxyacids lacking a P–C bond, ie, the esters. Salts and esters are described by naming the cations or organic group and changing the phosphorus suffix from *-ic* to *-ate* and *-ous* to *-ite*. Compounds are named according to the acids from which they are derived, eg, (C₂H₅O)₂P(=O)H is diethyl phosphonate [762-04-9], and (C₆H₅O)₃P is triphenyl phosphite [101-02-0]. Table 3 lists the structures and names of organo derivatives having P–C bond(s).

Table 1. Oxyacids of Phosphorus

Name and CAS Registry Number	Oxidation state	Molecular formula	Structure	Basicity and salts
(ortho)phosphoric acid[7664-38-2]	+5	H ₃ PO ₄		tribasic; salts are called phosphates polymeric forms, eg, H ₄ P ₂ O ₇ , H ₅ P ₃ O ₁₀ , (HPO ₃) _n , etc
hypophosphoric acid [7803-60-3]	+4, +4	H ₄ P ₂ O		tetrabasic; salts are called hypo-phosphates
isohypophosphoric acid[20698-54-8]	+3, +5	H ₄ P ₂ O		tribasic; salts are called isohypophosphates
phosphonic acid [13598-36-2]	+3	H ₃ PO ₃		dibasic; salts are called phosphites
diphosphonic acid [36465-90-4] (pyrophosphorous acid)	+3, +3	H ₄ P ₂ O ₅		dibasic; salts are called pyrophosphites
hypophosphonic acid [20267-10-1]	+2, +2	H ₄ P ₂ O ₄		monobasic; salts are called hypodi-phosphites
phosphinic acid [6303-21-5] (hypophosphorous acid)	+1	H ₃ PO ₂		monobasic; salts are called hypo-phosphites

Table 2. Organic Esters of Phosphorus Oxyacids

Phosphorus oxidation state	Molecular formula	Structure	Name	Parent oxyacid
----------------------------	-------------------	-----------	------	----------------

+5	$R_xH_{3-x}PO_4$	$\begin{array}{c} \text{O} \\ \parallel \\ \text{RO}-\text{P}-\text{OR}' \\ \\ \text{OR}'' \end{array}$	phosphate ester (phosphate)	(ortho)phosphoric acid, mono-, di-, and triesters exist
+5	$R_xH_{4-x}P_2O_7$	$\begin{array}{c} \text{O} \qquad \qquad \text{O} \\ \parallel \qquad \qquad \parallel \\ \text{RO}-\text{P}-\text{O}-\text{P}-\text{OR}' \\ \qquad \qquad \\ \text{OR}'' \qquad \text{OR}''' \end{array}$	diphosphate (pyrophosphate) ester	pyrophosphoric acid ^a
+3	R_2HPO_3	$\begin{array}{c} \text{O} \\ \parallel \\ \text{H}-\text{P}-\text{OR} \\ \\ \text{OR}' \end{array}$	phosphonate ester (phosphite)	phosphonic acid, mono- and diesters exist
+3	R_3PO_3	$\begin{array}{c} \text{O} \\ \parallel \\ \text{RO}-\text{P}-\text{OR} \\ \\ \text{OR} \end{array}$	triphosphite ester (phosphite)	hypothetical phosphorous acid, $P(OH)_3$; structure is isomeric with phosphonates
+1	RH_2PO_2	$\begin{array}{c} \text{O} \\ \parallel \\ \text{H}-\text{P}-\text{OR} \\ \\ \text{H} \end{array}$	phosphinite ester	phosphinic acid

^a Higher polymeric forms, eg, $(ROP)_n$, also exist (see PHOSPHORIC ACIDS AND PHOSPHATES).

Table 3. Organophosphorus Oxyacids and Derivatives

Phosphorus oxidation state	Acid formula	Acid structure	Name	Derivatives
+3	RH_2PO_3	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{P}-\text{OH} \\ \\ \text{OH} \end{array}$	organophosphonic acid (phos-phonates)	acid; mono- and diesters from phosphonic acid ^a
+1	RH_2PO_2	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{P}-\text{OH} \\ \\ \text{H} \end{array}$	organophosphinic acid; also phos-phonous acid (phosphinates)	acid; monoesters from phosphinic acid
+1	R_2HPO_2	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{P}-\text{OH} \\ \\ \text{R}' \end{array}$	organophosphinic acid; also phos-phonous acid (phosphinates)	acid; monoesters from phosphinic acid
1	R_2HPO	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{P}-\text{OH} \\ \\ \text{R}' \end{array}$	organophosphinous acid (phos-phinites)	acid; monoesters from hypothetical phosphinous acid, H_2POH ^b

^a Structure is isomeric with triphosphites.
^b Structure is isomeric with secondary phosphine oxides.

Other phosphorus compounds may also be considered as derivatives of the oxyacids. The P–O or P=O moieties may be replaced with isoelectronic groups to yield halo, P–X; amide, P(O)–NR₂; thio, P=S; imide, P=NH; etc. However, many phosphorus compounds are named as salts with phosphorus as the metallic or electropositive element, eg, phosphorus trichloride [7719-12-2], PCl₃; phosphorus oxybromide [7789-59-5] (phosphoryl bromide), POBr₃; phosphorus pentafluoride [7647-19-0], PF₅; phosphorus triamide [13566-19-3], P(NH₂)₃; and tetraphosphorus decasulfide [1314-80-3] (phosphorus pentasulfide; phosphorus(V) sulfide), P₄S₁₀. There is some latitude in describing halo derivatives of oxyacides. For example, diethyl phosphorochloridate [814-49-3], (C₂H₅O)₂P(O)Cl; methylphosphonic dichloride [676-97-1], CH₃P(O)Cl₂; phenylphosphinous dichloride [644-97-3] (phenyldichlorophosphine), C₆H₅PCL₂; and phenylthiophosphorodifluoridate [658-35-5], C₆H₅OP(S)F₂.

Some compounds are named as derivatives of the simple phosphorus hydrides (phosphines). For example, dimethylphosphine [676-59-5], (CH₃)₃PH; triphenylphosphine oxide [791-28-6], (C₆H₅)₃P=O; 1,2-dimethyldiphosphine [53684-00-7], CH₃PH₂PCH₃; diethyliodophosphine [20472-47-3], (C₂H₅)₂PI; phosphonium iodide [12125-09-6], PH⁺₄I[–]; tetramethylphosphonium chloride [1941-19-1], (CH₃)₄P⁺Cl[–]; and phenylphosphonium bromide [55671-96-0], C₆H₅PH⁺₃Br[–].

Chemical Properties

Oxidation States, Coordination Numbers, and Geometries. Phosphorus has electronic structure 1s²2s²2p 3s²3p³. There are thus five valence electrons, three of which are unpaired. Phosphorus bonding is primarily covalent because of its intermediate electronegativity, *X* = 2.1, which is equal to that of hydrogen. Oxidation states range from –3 to +5, but phosphorus exists in nature almost exclusively as phosphates or phosphate derivatives. The +5 oxidation state is the most stable for the oxide derivatives in both acidic and basic media as evidenced by the oxidation potentials (4).

<i>in 1 M acid</i>	$\text{PH}_3 \xrightarrow{+0.06 \text{ V}} \text{P} \xrightarrow{+0.51 \text{ V}} \text{H}_3\text{PO}_2 \xrightarrow{+0.50 \text{ V}} \text{H}_3\text{PO}_3 \xrightarrow{+0.28 \text{ V}} \text{H}_3\text{PO}_4$
<i>in 1 M base</i>	$\text{PH}_3 \xrightarrow{+0.89 \text{ V}} \text{P} \xrightarrow{+2.05 \text{ V}} \text{H}_2\text{PO}_2^- \xrightarrow{+1.56 \text{ V}} \text{HPO}_3^{2-} \xrightarrow{+1.12 \text{ V}} \text{PO}_4^{3-}$
<i>phosphorous oxidation state:</i>	IIIIIIIIIIV

The oxidation potentials also indicate the tendency for intermediate oxidation states to disproportionate.

The most common coordination numbers for the phosphorus atom are three, four, or five, although covalent linkages can range anywhere from one to six. The quadruply connected compounds appear to exhibit the highest stability, followed by the coordination numbers three and then five. Singly or doubly connected phosphorus compounds are typically unstable, and fewer examples of the five- and six-coordinate compounds are known. The phosphorus atom most commonly utilizes the *s* and *p* orbitals in σ-bonding. However, the promotional energy of 3*s* → 3*d* is small enough to allow the

vacant *d* orbitals to participate in hybridized orbital bonding for higher coordinate compounds. Bonding is almost exclusively limited to σ -bonds whatever the hybridization. The one notable exception is the partial π -character to the bonding in four-coordinate compounds.

The geometries of the phosphorus atom are related to the hybridization and the coordination number. Some of the more commonly encountered hybridizations and their corresponding spatial arrangements include the following.

Hybridization	Coordination number	Geometry
p^3	three	orthogonal trigonal
sp^3	four	tetrahedron
sp^3d	five	trigonal bipyramid
sp^3d^2	six	octahedron

Hybridization can be of mixed character and the bond angles in compounds can vary from ideal orientations of the pure hybrid.

The coordination number of four is the most common for phosphorus compounds containing the first and second row Group 13(III), 14(IV), and 15(V) elements. There are several reasons for the stability of the quadruply connected configuration. Lower coordination numbers do not offer as much total bonding energy, whereas higher coordination numbers involve repulsive steric interactions. Because phosphorus requires only three electrons to fill the valence shell, at least one of the four attached groups must be an electron acceptor. Atoms such as O and F contribute large bond energies. Unlike compounds of other coordination numbers, quadruply connected phosphorus compounds may include considerable π -bonding character. About one π -bond is present in each phosphorus atom, generated by electron donation (back bonding) from the bonded atoms into the empty phosphorus *d*-orbitals. The π -bonding character is often spread among the four σ -bonds. As a result, bond angles for quadruply connected compounds are typically 100–102°, somewhat smaller than the tetrahedral angle of 109°28'. In addition, π -bonding effectively reduces the ionic character of these compounds. Phosphorus compounds having a coordination number of three, five, or six exhibit no multiple or π -bonding character, except for minor amounts when highly electronegative constituents are present.

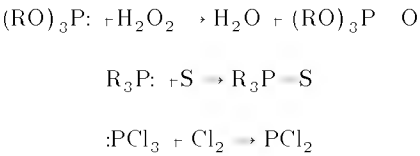
Pure tetrahedral coordination probably occurs only in species where there are four identical groups and no steric distortions. Both PCl_4^- and PBr_4^- , present in solid phosphorus halides, appear to have T_d point symmetry. Other species, eg, H_3PO_3 and POCl_3 , have only slightly distorted tetrahedra. Similar geometries occur in salts, esters, and other derivatives of phosphoric, phosphonic, and phosphinic acids as well as phosphine oxides and phosphonium salts.

The triply connected phosphorus compounds have a lone electron pair that dominates much of the chemistry for these compounds. Triply connected compounds typically exhibit pyramidal symmetry arising from p^3 hybridization. A considerable amount of sp^3 character may be present as well. Bond angles range near 100° vs 90° theoretical. Tricoordinate compounds typically act as electron donors, forming metal coordination compounds and addition compounds such as $\text{H}_3\text{P}\cdot\text{BF}_3$ [41593-56-0].

An unshared electron pair is characteristic of many tricoordinate phosphorus compounds. The availability of the electron pair for donation to an electron-pair acceptor depends on the character of the substituents on the phosphorus atom. Strongly electronegative groups, eg, chloro and fluoro, repress the availability of the unshared electron pair, and complexes that form between phosphorus trichloride or phosphorus trifluoride and an electron-pair acceptor are weak. However, alkyl groups can enhance the availability of an unshared electron pair. Trialkyl phosphines generally are potent electron donors and form stronger complexes than the corresponding amines. Intermediate between these, in order of decreasing availability of the associated electron pair, are aryl, hydrogen, alkoxy, and aryloxy phosphines.

The trigonal pyramid probably is the second most pervasive coordination pattern around phosphorus. Examples include phosphorus trihalides, phosphines, and derivatives of phosphorous and phosphinous acids. Formally, these can be considered distorted tetrahedra in each of which an unshared electron pair occupies the fourth coordination position. In fact, if the three substituents are different from one another, the tricoordinate compound exhibits optical activity such as that imparted by a tetracoordinate carbon atom having four different substituents. Most bond angles in trigonal pyramidal phosphorus compounds are less than 109° and between 90–100° for many phosphines (5). Such severe distortions from an idealized sp^3 hybrid are unlikely to result from electrostatic repulsion between the unshared electron pair and ions, eg, F⁻ and Cl⁻, especially in the case of phosphine, in which the ligands are more like protons than hydride ions. However, because the unshared pair occupies the spherical *s* orbital and the other ligands are bound to essentially pure orthogonal *p* orbitals (p^3 hybridization), electrostatic repulsions account for most observations. Thus, bond angles increase in the series as follows: PH_3 , 94°; PI_3 , 98°; PBr_3 , 100°; PCl_3 , 100.1°; and PF_3 , 104°.

An unshared electron pair on phosphorus reacts with oxidizing agents, eg, hydrogen peroxide, sulfur, or halogens,



and undergoes quaternization with alkyl halides,



Tri- and pentacoordinate phosphorus compounds often react by electron-pair mechanisms as demonstrated by the nucleophilic reactivity of the lone pair electrons in trivalent compounds, and the electrophilicity of the phosphorus atom in the pentavalent compounds. Some compounds also react by free-radical mechanisms. The theoretical and synthetic aspects of the chemistry of phosphorus compounds have been described (6–9).

The participation of phosphorus *d* orbitals in the five- and six-coordinate compounds provides increased polarizability, nucleophilicity, and ionic character. In fact, compounds such as phosphorus pentachloride [10026-13-8], PCl_5 , are thought to have considerable ionic character. The σ -bond orders between the phosphorus atom and its constituents in the higher coordinate compounds may be less than one.

The trigonal bipyramid and octahedron occur primarily in phosphorus pentahalides and the corresponding hexahalophosphate anions. Certain chelating diols form esters in which five or six groups are packed around phosphorus in these symmetries. The trigonal bipyramid should have three short bonds in the equatorial plane, ie, sp^3 hybrid orbital bonds, and two longer axial bonds that form from a hybrid of the remaining *p* orbital and one of the *d* orbitals. However, except at low temperatures, the fluorines in phosphorus pentafluoride [7647-19-0], PF_5 , are equivalent according to ¹⁹F-nmr measurements; this also is true for tetrafluoromethylphosphorus [420-64-4], CH_3PF_4 . Equivalence is attributed to pseudorotation in which the axial and equatorial groups exchange without breaking the bonds to phosphorus (10) via either a Berry pseudorotation (Fig. 1) in which two axial and two equatorial groups exchange by way of a square pyramidal transition state, or a turnstile pseudorotation in which two equatorial and one axial group exchange. If the frequency of interchange is rapid compared to time scale of the measurement, a distinction between axial and equatorial groups cannot be made.

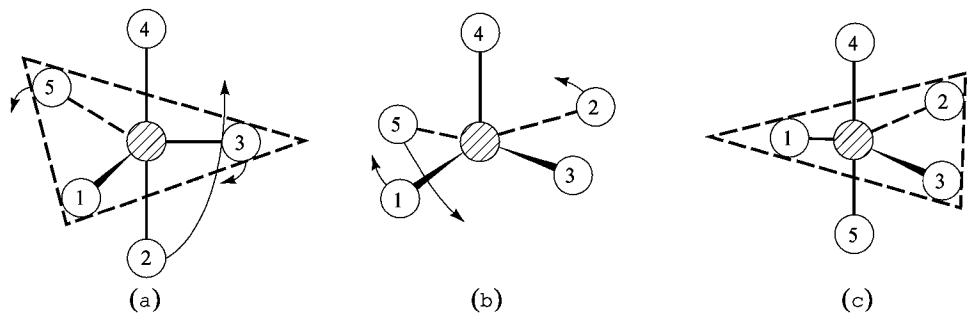


Fig. 1. Berry pseudorotation about pentacoordinate (⊗) phosphorus, where (○) represent fluorine atoms. (a) Original trigonal bipyramid; (b) square pyramidal intermediates; and (c) new trigonal bipyramid.

Much effort has been placed in the synthesis of compounds possessing a chiral center at the phosphorus atom, particularly three- and four-coordinate compounds such as tertiary phosphines, phosphine oxides, phosphonates, phosphinates, and phosphate esters (11). Some enantiomers are known to exhibit a variety of biological activities and are therefore of interest as agricultural chemicals, pharmaceuticals (qv), etc. Homochiral bisphosphines are commonly used in catalytic asymmetric syntheses providing good enantioselectivities (see also NUCLEIC ACIDS). Excellent reviews of low coordinate (coordination numbers 1 and 2) phosphorus compounds are available (12).

Bond Properties. Bond strengths, bond lengths, and atom electronegativity differences of various phosphorus–atom linkages are given in Table 4.

Table 4. Properties and Electronegativity Differences of Phosphorus–Atom Bonds^a

Bond, P–X	Bond energy, kJ mol ^b	Bond length, nm	X _X ^c –X _P
P–O	360	0.15–0.17	1.4
P=O	544	0.141–0.151	1.4
P–H	322	0.140–0.146	0.0
P–C	272	0.183–0.194	0.4
P–F	527	0.150–0.160	1.9
P–Cl	331	0.204–0.205	0.9
P–Br	264	0.215–0.220	0.7
P–I	184	0.248–0.252	0.4
P–N	230	0.17–0.18	0.9
P–P	209	0.217–0.227	0.0
P–S	230	0.200–0.215	0.4
P=S		0.187–0.197	0.4

^a Refs. 6 and 13.
^b To convert J to cal, divide by 4.184.
^c Pauling's electronegativity difference, where X_X represents the electronegativity of atom X.

Phosphorus–Oxygen and Phosphorus–Sulfur Bonds. According to valence bond (VB) theory, phosphorus–oxygen or phosphorus–sulfur linkages may be either single or double bonds. VB theory, a convenient accounting mechanism, may not accurately reflect the bond order or the nature of the bonding. In phosphoryl compounds, for example, the P=O bond may be considered to be formed via donation of the electron pair from the phosphorus atom to the oxygen atom, resulting in a bond of primarily σ-character. This bond may be more accurately depicted as a coordinate bond, P → O, or as P⁺–O[–], which indicates the ionic character of the covalent bond. The P=O description is customarily used in writing structures, yet this formalism does not necessarily imply π-bonding. In a species such as the phosphate ion, PO₄^{3–}, all phosphorus–oxygen bonds are equivalent, although the ion is commonly depicted as having one P=O. The presence of a P=O bond is associated with a high stability. Phosphorus has a strong affinity for oxygen, and the great tendency of formation of a P=O linkage dominates much of phosphorus chemistry.

The O or S atoms in P=O and P=S groups may act as electron donors although these groups form relatively weak complexes with electron acceptor compounds such as nonpolarizable, more electropositive (ie, hard) acids, including protons (14). Use is made of this property in the recovery of uranium from wet-process phosphoric acid by extractants such as trioctylphosphine oxide [78-50-2] and di(2-ethylhexyl) hydrogen phosphate [298-07-7].

The P=S linkages are typically more reactive than the relatively inert P=O linkages. P–S and P=S bonds are also less stable, hydrolytically and thermally, than their oxygen analogues. The P=S linkages are stronger electron donors. Therefore, complexes with electron acceptors are more stable if also more labile. Thus, they are useful as blocking groups for unshared electron pairs in some sensitive syntheses (see NUCLEIC ACIDS). The required manipulations are performed on the thio analogue,



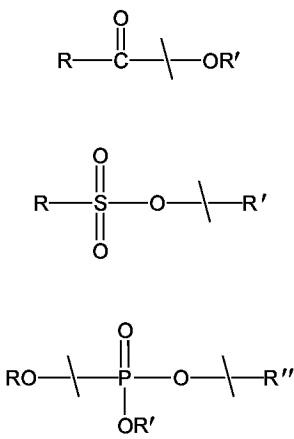
and the electron pair does not become involved with a reactant that can produce a phosphoryl group,



The thio group can be removed by a reagent that binds sulfur strongly, eg, triphenylphosphine [603-35-0].

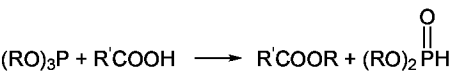
The behavior of the P–O linkage is in many respects intermediate between those of the C–O and S–O bonds. Oxo-acids of phosphorus are intermediate in acid strength between those of carbon and sulfur, and the strong acid function of phosphorus oxo-acids typically exhibit a pK_a near 2, eg, H₃PO₄ has a pK_a of 2.0; H₃PO₃, a pK_a of 1.5; and CH₃PO₃H₂, a pK_a of 2.3. Phosphorus esters are also more labile than carbonates or carboxylates but less so than sulfates. Although carboxylates hydrolyze by cleavage of the C–O bond and sulfates by cleavage of the O–R bond, phosphorus esters may cleave

at either point.



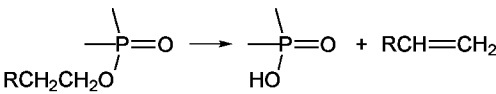
Like P–O–C linkages, P–O–P linkages are susceptible to hydrolytic degradation. Scrambling or interchange usually occurs for phosphorus oxyesters at temperatures and acidities lower than those required for the carbon esters but greater than those for the sulfur esters.

Trialkyl phosphites are good esterification reagents:

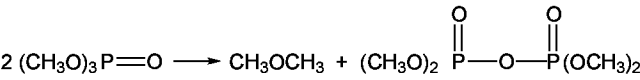


The yield of this reaction for R=R'=C₂H₅ is 73% (15). Reagents are less toxic than the corresponding sulfates and less volatile than the orthoformates.

Esters of phosphorus oxo-acids having beta hydrogens undergo olefin elimination upon pyrolysis, usually beginning at 160–200°C.

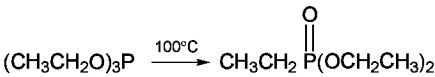


The resulting acid function catalyzes the same reaction. Therefore, it is desirable to distill phosphorus esters below 160°C and, if possible, in the presence of a mild acid scavenger. Phosphites and hydrogen phosphates are most susceptible to this decomposition, whereas phosphates and alkyl or aryl phosphonates are least susceptible. Ester groups lacking a beta hydrogen rearrange to form ether and anhydride under more severe conditions.



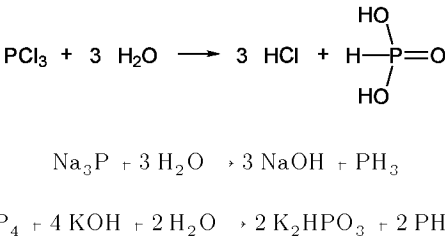
Tetraalkyl pyrophosphates are extremely toxic and caution should be used in the pyrolysis of such esters (16).

Trialkyl phosphites participate in Michaelis-Arbuzov rearrangement with alkyl halides. Such esters can undergo auto-Arbuzov rearrangement in the absence of an alkyl halide.



The exothermic reaction can be explosive. Therefore, it is preferable to initiate the reaction by using an alkyl halide at a lower temperature at which the reaction rate is controlled by the halide concentration.

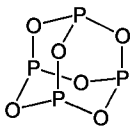
Phosphorus–Hydrogen Bond. A hydrogen bound to phosphorus has little acidic or hydric character. Most of the reactions the bond undergoes are those of a reducing agent. P–H bonds are formed by hydrolysis of active metal phosphides or phosphorus halides, by the rearrangement of P–O–H or P–S–H linkages, or by the hydrolysis of P–P bonds (6,17).



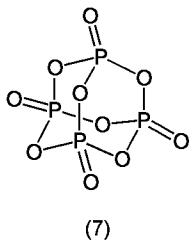
P–H bonds are readily susceptible to acid-, base-, or radical-catalyzed nucleophilic additions to unsaturated centers such as olefins and ketones, resulting in the concomitant formation of a P–C bond. The acid-catalyzed addition proceeds through the generation of an electron-deficient center that is attacked by the unshared pair on phosphorus. The base-catalyzed reaction is thought to proceed via formation of a negatively charged phosphide. Free-radical reactions of phosphine proceed through ready homolysis of P–H bonds.



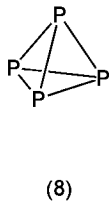
A P–H bond is readily attacked by active oxidizing agents.



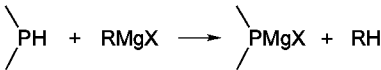
(6)



The chlorination of dialkyl phosphonates provides a convenient synthesis of phosphorochloridates:



Exclusion of light prevents attack of chlorine on the ethyl groups.
Phosphorus–hydrogen compounds undergo a metathetical exchange with some organometallic alkylating reagents (see GRIGNARD REACTIONS):



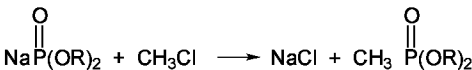
Some phosphorus–hydrogen compounds are pyrophoric, eg, diphosphine [13445-50-6], P_2H_4 , a common impurity in phosphine. Such contaminated phosphine usually ignites spontaneously on contact with air.
Phosphorus–Carbon Bond. The P–C bond is 0.184–0.194-nm long and has an energy of ca 272 kJ mol (65 kcal mol). It is one of the more stable bonds formed by phosphorus, resistant to both hydrolysis and oxidation (7,8). Unlike the phosphorus–halogen or phosphorus–oxygen bonds, the P–C linkage is inert to exchange. A phosphorus atom connected to carbon behaves similarly to another carbon atom in a hydrocarbon chain.
A general method of forming a phosphorus–carbon bond is by reaction of an organometallic reagent and a phosphorus–halogen link.



RM can be a traditional Grignard reagent or an organolithium, zinc, aluminum, or mercury compound. The Grignard route is employed commercially for production of tertiary phosphines, even though these reactions are subject to side reactions. Yields are often low, eg, 40–50% for $(\text{C}_4\text{H}_9)_3\text{P}$ prepared via a Grignard reaction (18). A phosphorus–carbon bond can form from the metathetical reaction of a phosphorus halide and a pseudohalide salt.



In general, compounds having an active phosphorus–metal linkage react with alkyl halides. Such compounds include alkali or alkaline-earth phosphides or phosphine derivatives, eg, Na_3P , PH_2Na , XMgPR_2 , or

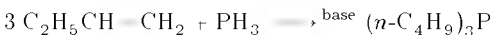
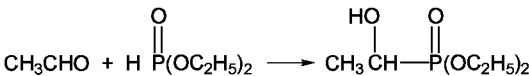


Friedel-Crafts reaction, utilizing phosphorus trichloride or phosphorus tribromide [7789-60-8] in place of an acyl halide, can also be used for the preparation of many arylphosphorus compounds (8) (see FRIEDEL CRAFTS REACTIONS).

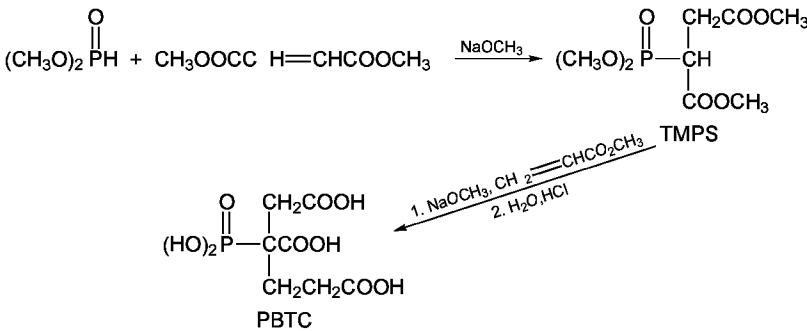


Ar represents an aryl group. Diaryl products are obtained after long reaction times. Other Friedel-Crafts catalysts, eg, ZnCl_2 , FeCl_2 , HF, and BF_3 , can also be used. In most cases, stoichiometric amounts of the catalyst are required. However, strong complexation of the phosphine by the catalyst necessitates separation by vacuum distillation, hydrolysis, or addition of reagents such as POCl_3 to form more stable aluminum chloride complexes. Whereas yields up to 70–80% are possible for some aryl derivatives, yields of aliphatic derivatives are generally much less (19).

An important feature of P–H bonds is the ready addition across unsaturated groups such as aldehydes or olefins to form P–C bonds.

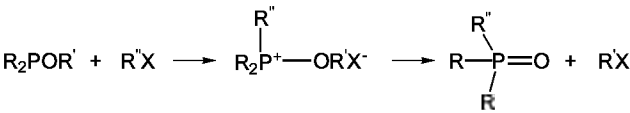


Another important P–C-bond-forming reaction is the base-catalyzed Michael addition to activated double bonds. For example, dimethyl phosphite can be added to dimethyl maleate to yield tetramethylphosphonosuccinate [2788-26-3] (TMPS), an intermediate in the synthesis of 2-phosphonobutane-1,2,4-tricarboxylic acid [37971-36-1] (PBTC) with 98% yield (20).

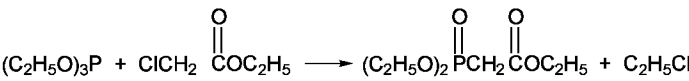


PBTC is a water treatment chemical sold by Bayer under the trade name Bayhibit AM. The addition reactions can also be operated as a continuous process (21).

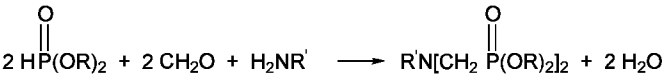
One of the most useful reactions in forming a P–C bond is the Michaelis-Arbusov reaction, which is a characteristic reaction of tricoordinate phosphorus compounds containing an alkoxy group (22). Alkylation of the electron pair is followed by rearrangement of the initial phosphonium salt.



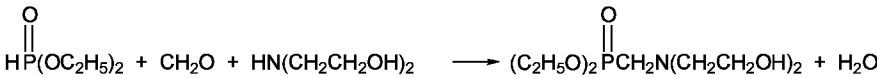
The reaction proceeds through an intermediate phosphonium salt which can be isolated in some instances. The Michaelis-Arbusov reaction is especially useful for converting trialkyl phosphites, (RO)₃P, to alkylphosphonic esters, and to esters of phosphonocarboxylic acids.



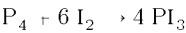
Triethylphosphonoacetic acid [867-13-0] (TEPA) is a useful olefination reagent for Homer-Emmons reactions in organic synthesis. A modified Mannich reaction is useful for the preparation of aminomethylphosphonates used as water treatment chemicals (23).



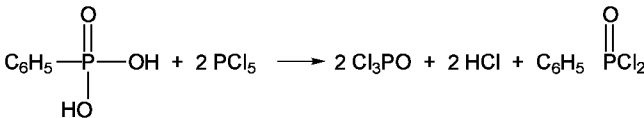
This reaction is catalyzed by hydrogen chloride and yields can be essentially quantitative when using either free phosphonic acid or its diesters. The flame retardant, Fyrol 6, produced by Akzo Chemicals, Inc. and used for rigid urethane foams, is synthesized as follows (24).



Phosphorus–Halogen Bonds. The bond length and bond energy of phosphorus–halogen compound are presented in Table 4. Bond lengths increase and bond energies decrease with increasing ionic radii. Phosphorus–halogen bonds are formed by the action of the elements, eg,

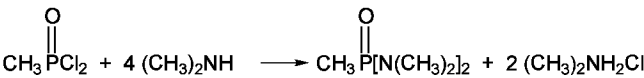


halogenation of P–H bonds, and the action of common halogenating agents on acids and esters of phosphorus. For example, phenylphosphonic acid [1571-33-1] reacts with phosphorus pentachloride to form phenylphosphonic dichloride [824-72-6].

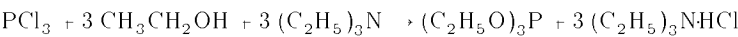
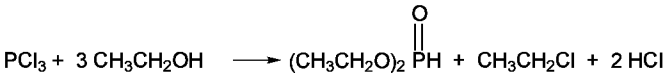


Despite the large energy of formation of many phosphorus–halogen bonds, these are among the most labile phosphorus bonds and participate readily in metathetical reactions. Mixtures of phosphorus halides scramble the halogens to yield mixed halide combinations (6). Thus, mixtures of POCl₃ and phosphorus oxybromide [7789-59-5], POBr₃, contain not only these two pure materials but also POCl₂Br [13455-03-3] and POClBr₂ [13550-31-7]. Other ligands on phosphorus, eg, ester or amide groups, also scramble but may require thermal excitation for the process.

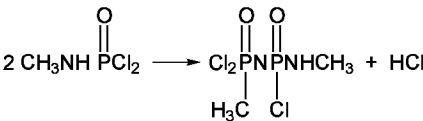
Phosphorus halides are subject to reactions with active hydrogen compounds and result in the elimination of hydrogen halide. They are convenient reagents in the synthesis of many esters, amides, and related compounds. However, because the involved hydrogen halide frequently catalyzes side reactions, it is usually necessary to employ a hydrogen halide scavenger to remove the by-product.



In some instances, the presence of the acid scavenger affects the nature of the product.



When a phosphorus–halogen bond and an active hydrogen occur in the same molecule, such species are subject to thermal rearrangement.



The rate of reaction of a phosphorus halogen and an active hydrogen depends on a number of factors, including steric requirements of the reactants, inductive effects on the phosphorus, and size of the halogen, eg, iodine is the most active and fluorine is the least active. Thus, PCl₅ is a highly active phosphorus chloride whereas the chlorocyclophosphazenes are relatively inactive (25). The nearly planar P₃N₃Cl₃ is quite inactive but the highly puckered P₄N₄Cl₈ [2950-45-9] hydrolyzes slowly on exposure to atmospheric moisture.

Phosphorus Sulfides

Properties and Reactions. Phosphorus combines with sulfur to form the binary tetraphosphorus trisulfide [1314-85-8] (phosphorus sesquisulfide), P₄S₃, (1); tetraphosphorus pentasulfide [12137-70-1], P₄S₅, (2); tetraphosphorus heptasulfide [12037-82-0], P₄S₇, (3); and phosphorus(V) sulfide [1314-80-3] (tetraphosphorus decasulfide), P₄S₁₀, (4). Further, tetraphosphorus enneasulfide [25070-46-6], P₄S₉, (5) has also been reported (26). In addition, a stable oxysulfide, P₄O S₄ [15780-31-1], exists as a colorless, deliquescent crystalline solid which has a melting point of 102°C. Some physical constants and thermodynamic data for these compounds are presented in Table 5. The structures, determined by x-ray crystallography, are shown in Figure 2 (26).

Table 5. Properties of Phosphorus Sulfides^a

Property	P ₄ S ₃	P ₄ S ₅	P ₄ S ₇	P ₄ S ₉	P ₄ S ₁₀
CAS Registry Number	[1314-85-8]	[12137-70-1]	[12037-82-0]	[25070-46-6]	[1314-80-3]
mp, °C	173	170–220	307	240–270	285
bp, °C	407	dec ^b	523		515
density, g mL	2.03	2.17	2.19	2.08	2.09
color of solid	yellow	light yellow	almost white		yellow
solubility in CS ₂ , g 100 g					
20° C	11				0.082
0° C	27		0.005		0.18
17° C	100 ^c	10	0.029		0.22
reactivity					
in cold water	little attack		readily dec	readily dec	slowly dec
in air	slowly oxidize		dec		slowly dec
bond length, pm ^d					
P—P	223	225	233	233	
P—S	209	212	211	212	211
P=S		194	192	192	192
ΔH _f ^e , kJ mol ^f	154	207	272	292	309
ΔH _{melt} ^g , kJ mol ^f					41
ΔH _{vap} ^g , kJ mol ^f	63.6		93.7		178 ^g

^a Ref. 26.
^b Decomposes to P₄S₃ and P₄S₇.
^c The solubility in C H at 17°C is 2.5 g 100 g; at 80°C, 17 g 100 g.
^d Values are given are average ones.
^e From red phosphorus and sulfur.
^f To convert J to cal, divide by 4.184.
^g Per mole of P₄S₁₀ equivalent; the vapor species is P₂S₅.

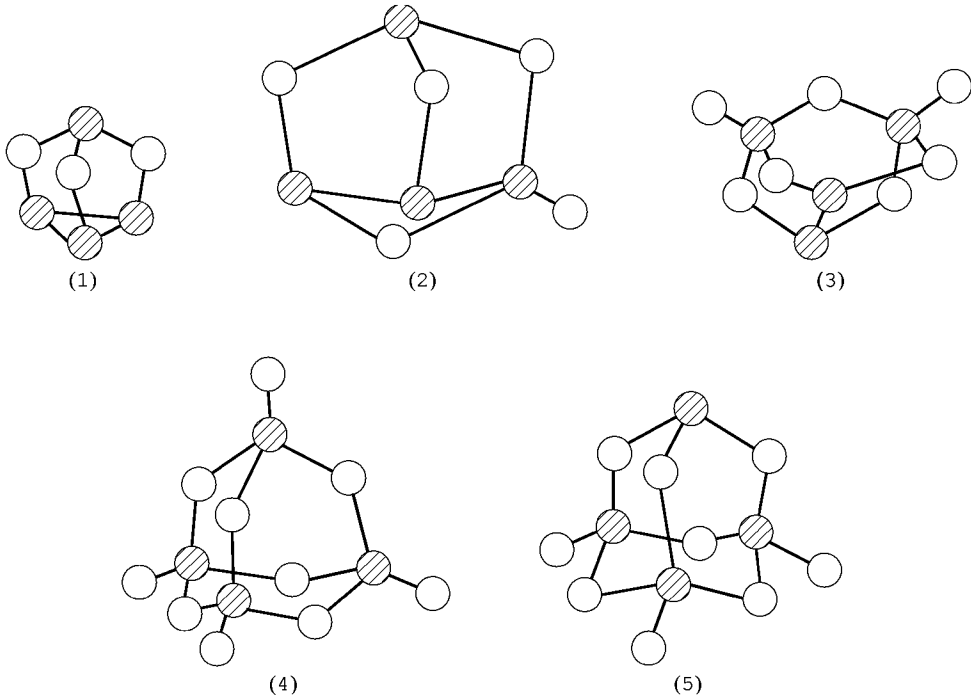


Fig. 2. Structure of phosphorus sulfides, where (⊗) represents phosphorus and (○) sulfur. See text.

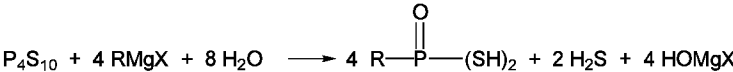
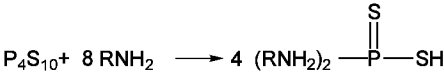
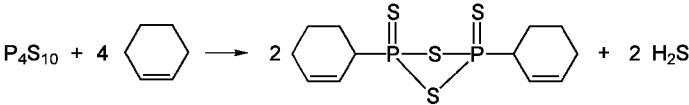
The hydrolysis of phosphorus sulfides has been studied quantitatively. A number of products are formed (Table 6). Whereas phosphorus(V) sulfide reacts slowly with cold water, the reaction is more rapid upon heating, producing mainly hydrogen sulfide and orthophosphoric acid, H₃PO₄. At high pH, P₄S₁₀ hydrolyzes to a mixture of products containing thiophosphates and sulfides.

Table 6. Products from Hydrolysis of Phosphorus Sulfides, %^a

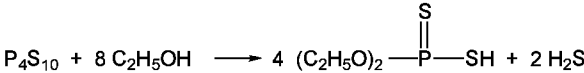
Product	Alkaline solutions			Acidic solutions			Neutral solutions		
	P ₄ S ₃	P ₄ S ₇	P ₄ S ₁₀	P ₄ S ₃	P ₄ S ₇	P ₄ S ₁₀	P ₄ S ₃	P ₄ S ₇	P ₄ S ₁₀
PH ₃	5	3			3		3	0	
H ₃ PO ₂	15	2	10		1	10	38	24	
H ₃ PO ₃	75	38			39	0	49	49	
H ₃ PO ₄	0	57	80		57	85	6	6	100

^a Ref. 6.

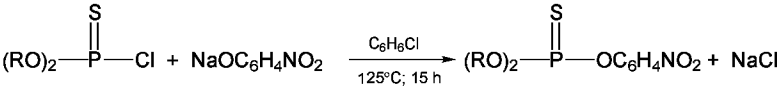
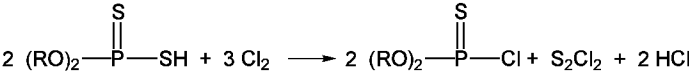
Phosphorus(V) sulfide reacts with olefins, amines, Grignard reagents, and terpenes (6,26) as follows:



O,O'-Dialkyl or diaryl dithiophosphoric acids are obtained readily from alcoholysis of phosphorus(V) sulfide. Thus, P₄S₁₀ reacts with ethyl alcohol as follows:



Dialkyl and diaryl dithiophosphoric acids are the bases of many high pressure lubricants, oil additives (see LUBRICATION AND LUBRICANTS), and ore flotation chemicals (see MINERAL RECOVERY AND PROCESSING). Organophosphorus insecticides such as Parathion are made by chlorination of the appropriate dialkyl dithiophosphate and subsequent reaction of the intermediate dialkyl thiophosphoric chloride with sodium *p*-nitrophenolate according to the following (see INSECT CONTROL TECHNOLOGY).



Manufacture. Phosphorus sulfides are manufactured commercially by direct reaction of the elements. Elemental phosphorus and sulfur are measured into a reaction vessel containing a heel of molten phosphorus sulfide. The reaction can be batch or continuous. The ratio of phosphorus to sulfur in the feed determines which phosphorus sulfur compound (Table 5) is formed. The reaction temperature can be the boiling point or lower. For the boiling reactor (27,28), the phosphorus sulfide product is first purified by distillation and then condensed to a liquid. Alternatively, the liquid product can be formed directly in a nondistilled process (29–31), which may involve a subsequent distillation step (30), and in which the phosphorus is often cleaned up prior to use (30–32). For either process, the liquid phosphorus sulfide product is solidified, and usually sized to form a commercial material.

Phosphorus(V) sulfide, an important commodity in the United States since about 1920, is the dominant commercial material. Phosphorus sesquisulfide, P₄S₃, has been made commercially since about 1900. Phosphorus heptasulfide was introduced as a small-scale commercial product in 1940.

Analyses and Analytical Test Methods for P₄S₁₀. A typical analysis of P₄S₁₀ yields the following.

Property	Description and value
appearance	greenish yellow powder
phosphorus, wt %	27.4–28.3
color in ethanol, APHA	20–60
iron, ppm	10–20
reactivity, °C min	
reactive grade	1
high reactive grade	4–10

Phosphorus content usually is measured by a double end point titration method in which a 1.0-g sample is dissolved in a hot HNO₃–H₂SO₄–HClO₄ mixture. The pH is adjusted to 2.5 with NaOH, and the resulting H₃PO₄ is titrated with 0.5-N NaOH, using an automatic titrator. The titer between the first and second end points is used to quantify the phosphorus as H₃PO₄.

Reactivity is measured by placing a standard quantity, 100 mL, of isopropyl alcohol in a 500- or 1000-mL Dewar flask equipped with a stirrer and a temperature-measuring device. The temperature of the alcohol is adjusted to 30°C. Thirty-six grams of the sample are added and the temperature is observed as a function of time from the addition until a maximum is reached. Reactivity is defined as the temperature rise divided by the time interval to reach this maximum. Other alcohols may also be used for measuring reactivity (30).

Reactivity is affected by particle size. Smaller particles react faster. However, the dominant factor for reactivity is the solidification rate. Material that is solidified quicker reacts faster with alcohol (30). Commercial P₄S₁₀ is a solid solution containing P₄S₁₀, P₄S₉, P₄S₇, free sulfur, etc (33). Solidification rate

also affects the distribution of these compounds. In part because P₄S₉ reacts 10 times faster with alcohol than P₄S₁₀ (34). Further, for commercial P₄S₁₀, the P₄S₉/P₄S₁₀ ratio decreases (33) and the reactivity decreases on annealing (28,30).

Shipping and Storage. Phosphorus(V) sulfide is stored and shipped in 208-L (55-gal) drums containing 250 kg of product, portable closed aluminum bins containing 1800–3400-kg net weight, and railcars. P₄S₁₀ is classified as a flammable solid having the international shipping code of UN No. 1340.

Health and Safety Factors, Toxicology. One source of danger in the handling of phosphorus(V) sulfide is hydrogen sulfide. The OSHA exposure limit to hydrogen sulfide gas, which has a rotten egg odor at 0.1 ppm, is 15 ppm for 15 minutes or 10 ppm for 8 hours. Respiratory problems occur for exposure above 50 ppm; death on exposure to concentration above 1000 ppm. Hydrogen sulfide results from the reaction of the phosphorus(V) sulfide with water or some other chemicals, such as alcohols. Further, phosphorus(V) sulfide combines with atmospheric moisture to release hydrogen sulfide via gradual hydrolysis. Care should be taken (1) to store P₂S₅ in well-sealed containers; (2) to convey P₂S₅ in well-sealed materials-handling systems containing dry, inert atmospheres; and (3) to handle phosphorus(V) sulfide in well-ventilated areas. Another source of danger in handling phosphorus(V) sulfide is fire and explosion. Phosphorus(V) sulfide burns readily to form sulfur dioxide, SO₂, and phosphorus pentoxide, P₄O₁₀, thus P₄S₁₀ should be handled in inert atmospheres. Ignition sources such as spark, static electricity, and heat should also be eliminated.

Phosphorus(V) sulfide is a mild skin irritant and may cause dermatitis in sensitive individuals. The primary health hazard results from the liberation of hydrogen sulfide after contact with moisture. Contact with moisture also forms phosphoric acid. A secondary hazard is the formation of sulfur dioxide when phosphorus(V) sulfide burns. The oral LD₅₀ of in rats is 389 mg/kg; the OSHA standard time-weighted average (TWA) is 1 mg/m³ (35).

Uses. Phosphorus(V) sulfide is used in the manufacture of lubricating oil additives, insecticides, ore flotation agents, and specialty chemicals. Phosphorus sesquisulfide, P₄S₃, has been used extensively in the manufacture of strikeanywhere matches (qv). In addition, small quantities are used in fireworks (see PYROTECHNICS).

Phosphorus Halides

Phosphorus forms well-defined halogen compounds of the types PX₃, PX₅, POX₃, and PSX₃, all of which except the pentaiodide and the oxy- and sulfoiodides are known. In addition to the binary halides, a few of the many possible mixed halides, eg, PX₂Y and PX₃Y₂, have been prepared. The commercially important phosphorus halides are phosphorus trichloride [7719-12-2], phosphorus oxychloride [10025-87-3], phosphorus pentachloride [10026-13-8], and phosphorus sulfochloride [3982-91-0]. A few other phosphorus halides, eg, PI₃, PBr₃, PBr₅, PF₃, and PF₅, are marketed as reagent chemicals.

The trihalides of phosphorus usually are obtained by direct halogenation under controlled conditions, eg, in carbon disulfide solution in the case of the triiodide. Phosphorus trifluoride [7647-19-0] is best made by transhalogenation of PCl₃ using AsF₃ or CaF₂. All of the phosphorus trihalides are both Lewis bases and acids. The phosphorus trihalides rapidly hydrolyze in water and are volatile. Examination by electron diffraction has confirmed pyramidal structures for the gaseous trihalide molecules (36). Physical properties and heat of formation of some phosphorus halides are listed in Table 7.

Table 7. Physical Properties of Phosphorus Halides^a

Compound	CAS Registry Number	Melting point, °C	Boiling point, °C	Specific gravity ^b	Physical state at STP ^c	Δ <i>H</i> _f , kJ/mol ^d	Δ <i>H</i> _{vap} , kJ/mol ^d	Critical temperature, °C
PCl ₃	[7719-12-2]	93.6	76.1	1.575	liquid	319.7	30.5	285.5
PCl ₅	[10026-13-8]	167 ^e	159 subl	2.114	solid ^f	443.5	64.9	372
POCl ₃	[10025-87-3]	+1.2	106.5	1.68	liquid	597.1	34.7	331.8
PSCl ₃	[3982-91-0]	36	125	1.668	liquid			
PBr ₃	[7789-60-0]	41.5	173.3	2.880	liquid	184.5	38.9	
PBr ₅	[7789-69-7]	83.8 (subl)	>106 dec		solid ^g	−269.9		
PI ₃	[13455-01-1]	61.2	120 dec		solid ^h	−45.6		
P ₂ I ₄	[13455-00-0]	125.5	dec		solid ⁱ			
PF ₃	[7783-55-3]	151.5	101	3.907	gas ^j	−918.8	14.6	2.05
PF ₅	[7647-19-0]	93.7	84.5	5.84	gas	1595.8	16.74	
POF ₃	[13478-20-1]	39.4	39.7	4.65	gas	1211.3	21.2	73.3
PSF ₃	[2404-52-6]	148.8	52.3		gas		19.6	72.8
PF ₂ Cl	[14335-40-1]	164.8	47.3		gas		17.6	89.17
PF ₃ Cl ₂	[13454-99-1]	124	2.5		gas		26.7	
PF ₄ Cl	[13637-88-2]	132	43.4				21.59	
PF ₂ Br	[15597-40-7]	133.8	16.1		gas		23.9	113
PF ₂ Br ₃	[13445-58-4]	20	~106		liquid			
PFCl ₂	[15597-63-4]	144.1	13.9		gas		24.89	
PFBr ₂	[15597-39-4]	145	78.4		liquid		31.88	254
PF ₂ I	[13819-11-9]	93.8	267					

^a Refs. 5, 36–39.
^b At 298.15 K and 101.3 kPa (1 atm).
^c Colorless unless otherwise noted.
^d To convert J to cal, divide by 4.184.
^e At 122 kPa (919 mm Hg).
^f White to pale yellow.
^g Red-brown.
^h Yellow.
ⁱ Dark red.
^j Ligh orange.

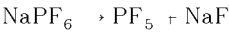
The melting, boiling, and sublimation points of many of the phosphorus halides are well defined and therefore serve for identification. Distillation is the easiest method of purification. Phosphorus-31 nmr can be used to analyze mixtures of halides that undergo halogen-exchange reactions.

The pentahalides of phosphorus, PX_5 , in the gas phase exhibit varying tendencies to dissociate into trihalide and halogen. Instability increases with increasing ionic radius of the halogen. The pentafluoride appears to be thermally stable. Dissociation of the pentachloride, a few percent at 100°C and 101.3 kPa (1 atm), is essentially completed at 300°C (36). The pentabromide is partially dissociated in the liquid state and totally dissociated above ca 35°C (39). Pentaiodide does not exist. The molecules of PF_5 and PCl_5 in the vapor phase are trigonal bipyramids. In the crystalline state, both pentachloride and pentabromide have ionic structures, ie, $[PCl_4]^+ [PCl]^-$ and $[PBr_4]^+ [PBr]^-$, respectively. The PX_4^+ cations are tetrahedral and the PX_6^- anion is octahedral (36,37).

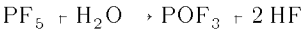
Phosphorus pentafluoride is synthesized easily by transhalogenation of phosphorus pentachloride using arsenic trifluoride at low temperature,



by heating hexafluorophosphate of sodium, calcium, and barium, and by



heating phosphorus pentachloride in the presence of the fluoride of a divalent metal, ie, Cd, Ba, Zn, or Pb, at 300–400°C. Phosphorus pentafluoride, a colorless gas that fumes in air and decomposes in water, is a strong Lewis acid.

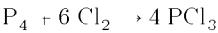


PF_5 forms complexes with amines, ethers, and other bases as well as F, with which phosphorus becomes six-coordinate. Dry phosphorus pentafluoride does not attack glass. The yellow crystalline phosphorus pentabromide forms from the reaction of PBr_3 and excess bromine.

Phosphorus halides and metals or metal salts form addition complexes. Some typical compounds are $PCl_5 \cdot SbCl_3$ and $PCl_5 \cdot AlCl_3$. The trivalent complexes contain metal–phosphorus bonds. The pentavalent complexes involve rearrangements to produce assemblies of tetrahedral PX_4^+ cations and various anions.

Phosphorus Trichloride.

Properties and Reactions. Phosphorus trichloride can be prepared either by direct chlorination of elemental phosphorus,



by reduction of $POCl_3$ by CO,

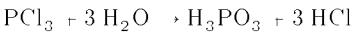


or by chlorination of ferrophosphide,

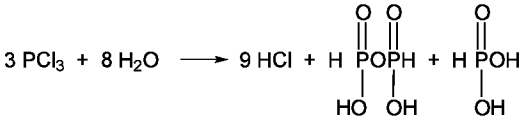


Phosphorus trichloride, PCl_3 , is a clear, volatile liquid having a pungent, irritating odor. Principal reactions of PCl_3 , an excellent chlorination reagent for various hydrocarbons, are summarized in Figure 3.

The reaction of phosphorus trichloride and water is highly exothermic and vigorous. Depending on the mole ratio of $H_2O : PCl_3$, three different products can result from hydrolysis. If the ratio is greater than 3, phosphonic acid is produced:



If the ratio is 2.5–3, the product is a mixture of phosphonic and pyrophosphonic acids:



When the $H_2O : PCl_3$ ratio falls below 2.5, a third product of indefinite composition, called lower oxides of phosphorus, is formed. The material ranges from a clear viscous polymer to yellow-to-orange solid flakes of limited solubility in PCl_3 ; ^{31}P -nmr identifies a polymer of randomly disposed P–P and P–O–P linkages. Other groups along the chain consist of –OH, –O, H, and sometimes –Cl or other halides (40). When this compound contacts water, halide groups hydrolyze more rapidly than the rest of the polymer, which leads to the formation of a mixture of phosphinic, phosphonic, and phosphoric acids and the evolution of phosphine and diphosphine gases. The gas evolution can proceed rapidly in the absence of water at elevated (>150°C) temperatures. Phosphine, when contaminated with only traces of diphosphine, ignites spontaneously in air. Controlled oxidations by hypochlorite or hydrogen peroxide can consume the yellow-orange solid flakes (41).

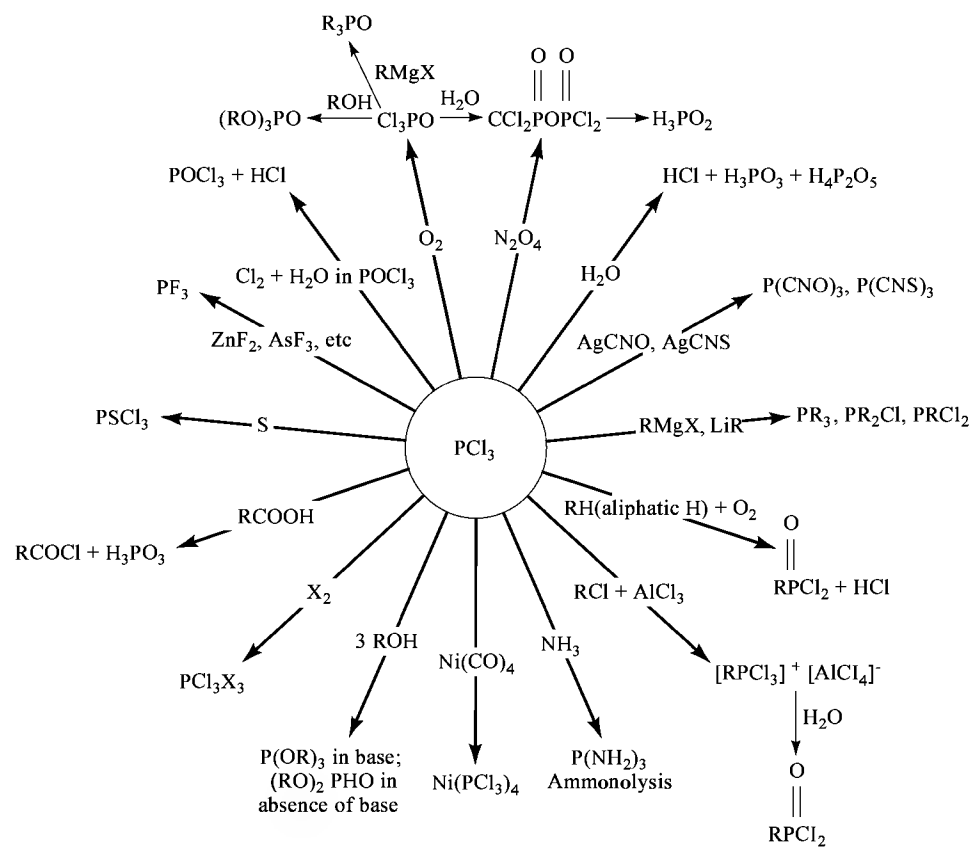


Fig. 3. Important reactions of PCl₃, where X is halogen and R is alkyl. Many of these reactions are typical for other PX₃ as well as POX₃ compounds.

Although PCl₃ is nearly insoluble in water, it hydrolyzes rapidly. Unless the addition of the chloride to water is carefully controlled by vigorous agitation, the water concentration in the vicinity of the interface between the PCl₃ and the aqueous solution may be depleted to the point where the ratio falls below 2.5 and formation of the polymer results. The polymer also forms during storage when PCl₃ is in contact with moist air. Such degradation is a problem in tank cars, storage tanks, and other reusable PCl₃ containers. The polymer, which is suspended in PCl₃, can be dissolved by the addition small quantities of PCl₅ or by Cl₂ gas chlorination resulting in the formation of POCl₃ and more PCl₃.

Manufacture. Phosphorus trichloride is made by direct union of the elements. The reaction is moderated by combining the chlorine and phosphorus in the presence of a precharge of phosphorus trichloride that is refluxed continuously. A typical manufacturing scheme is shown in Figure 4. Liquid phosphorus and chlorine gas are continuously introduced into the reaction vessel, which is arranged so that a significant portion of the phosphorus trichloride contained in it undergoes reflux. The remaining phosphorus trichloride is distilled into a pot. A yield of ca 99.0% or higher of purified phosphorus trichloride usually is obtained with respect to both the phosphorus and chlorine. The effect of raw material impurities on reactor operations and process waste generations has been evaluated (42). Most raw material impurities remain in the reactor and are removed periodically. Process safety can be improved by using an on-line analyzer and other procedures (43–45). Monsanto has the world's largest production unit for PCl₃.

Specifications and Analytical Methods. Typical analyses for technical-grade PCl₃ are as follows (46):

Property	Description or value
appearance	water-white liquid
distillation range, °C	
first drop	74.8
1–96 mL	0.9
dry point	76.0
specific gravity, 15.5°C 15.5°C	1.586
assay, %	99.75
Fe, ppm	1

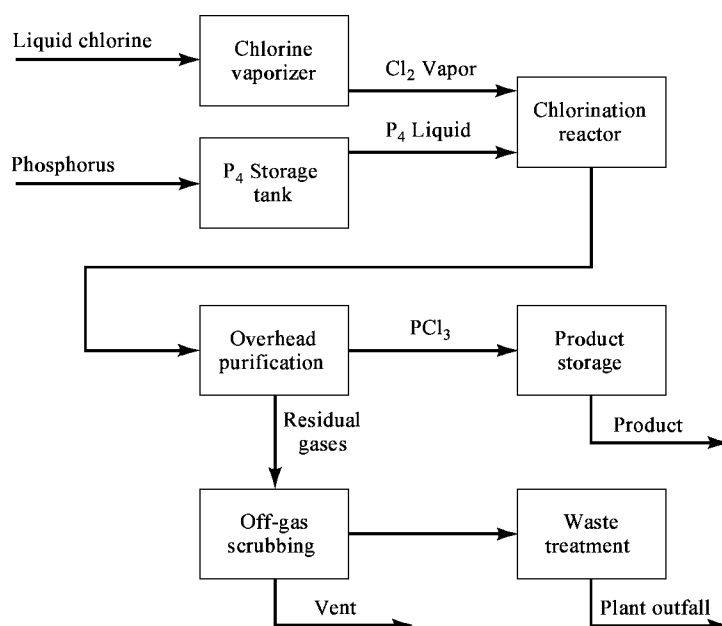


Fig. 4. Manufacturing of phosphorus trichloride.

Phosphorus oxychloride content and impurities are determined by gas chromatography analyses.

Storage, Shipping, and Handling. Phosphorus trichloride is classified by the ICC as a corrosive liquid and poison inhalation hazard. U.S. Department of Transportation (DOT) white acid label and red poison label are required by law on individual containers: DOT UN No. 1809 (46,47). Alloy or glass-lined vessels are used for storage and reactors. One-gallon (3.8-L) quantities of PCl_3 are stored and shipped in glass containers using specified wooden overpacking (46). Heresite-lined steel drums are for 208-L (55-gal) volume. Bulk PCl_3 shipments are made in Heresite-lined tank cars of 15,000–30,000 L (4000–8000 gal) each, and tank trucks of 15,000–19,000-L (4000–5000-gal) volume. The industrial trend is toward closed-loop loading–unloading operations.

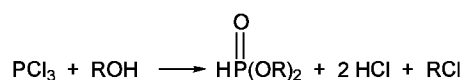
Health and Safety Factors, Toxicology. Phosphorus trichloride severely burns skin, eyes, and mucous membranes. Contaminated clothing must be removed immediately. Vapors from minor inhalation exposure can cause delayed onset of severe respiratory symptoms after 2–24 h, depending on the degree of exposure. Delayed, massive, or acute pulmonary edema and death can develop as consequences of inhalation exposure.

Phosphorus trichloride is highly toxic by ingestion and slightly toxic by single dermal applications. It reacts violently with water and can generate gases sufficient to cause rupture of closed or inadequately vented containers. If the gases contain diphosphine, they can ignite spontaneously. Acids produced in reactions of PCl_3 with water can evolve hydrogen gas on contact with metals. The threshold limit value (TLV) is 0.5 ppm or 3 mg m^{-3} . The oral LD_{50} in rats is 550 mg/kg; the inhalation LC_{50} for rats is 104 ppm/4 h, for guinea pigs, 50 ppm/4 h. The OSHA standard TWA is 0.5 ppm (35,46).

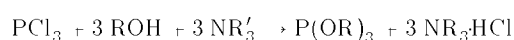
Uses. The largest usage of PCl_3 is to produce phosphonic acid, H_3PO_3 , which in reaction with iminodiacetic acid and formaldehyde forms a glyphosate intermediate that is decarboxymethylated to glyphosate, an effective nonselective herbicide (see HERBICIDES). Phosphorus trichloride is also a convenient chlorinating reagent for producing various acyl and alkyl chlorides.



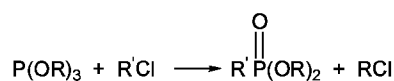
Phosphorus trichloride reacts readily with oxygen, sulfur, chlorine, and water. It serves as an intermediate in the production of phosphorus oxychloride, phosphorus sulfochloride, phosphorus pentachloride, and phosphonic (phosphorous) acids. PCl_3 is also the raw material for the manufacture of dialkyl phosphonates,



trialkylphosphites,



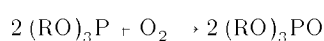
and dialkyl alkylphosphonates.



Alkyl or aryl phosphonates, which contain a carbon–phosphorus bond, are comparatively more stable. They are of interest as antiscaling additives and corrosion inhibitors for cooling towers and heat exchangers (see DISPERSANTS; WATER, INDUSTRIAL WATER TREATMENT), surfactants (qv), sequestrants, and textile-treating agents. Trialkyl phosphites are useful as esterification (qv) reagents.

The phosphonate esters, $\text{HP}(\text{=O}(\text{OR})_2)$, of alkylated phenols are used extensively as lubricating-oil additives to control bearing corrosion and oxidation, and to impart antirust properties as stabilizers, as antioxidants (qv) and flame retardants in plastics, as specialty solvents, and as intermediates (see CORROSION AND CORROSION CONTROL; HEAT STABILIZERS).

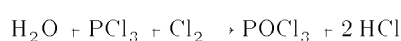
Phosphite triesters, $\text{P}(\text{OR})_3$, form donor complexes with transition metals and other acceptors and are oxidized to the respective phosphates under appropriate conditions.



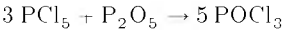
Most phosphate esters are made from POCl_3 and used as plasticizers (qv) and nonflammable hydraulic fluids (qv).

Phosphorus Oxychloride.

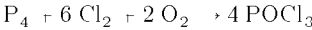
Properties and Reactions. Several methods of preparation are available for POCl_3 , including partial hydrolysis of PCl_5 by heating in the presence of oxalic or boric acid, chlorination and hydrolysis of PCl_3 in the presence of H_3PO_4 ,



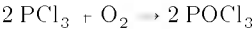
heating a mixture of P₂O₅ and PCl₅,



controlled oxidation and chlorination of elemental phosphorus,



oxidation of PCl₃ using ozone (qv) or oxygen (qv),



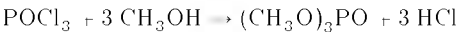
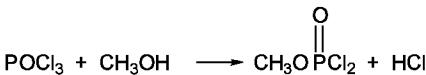
and heating calcium phosphate [7757-93-9] in a mixture of chlorine and carbon monoxide.



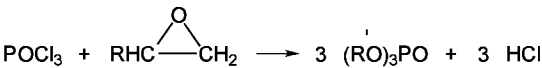
Phosphorus oxychloride [10021-87-3] (phosphoryl chloride), POCl₃, is a colorless fuming liquid having a pungent, disagreeable odor and is reactive with water. The reactions of POCl₃ resemble those of PCl₃. By means of Grignard reagents, the halogens can be replaced by alkyl or aryl groups.



The reaction proceeds in a stepwise fashion and products with only one or two halogens can be produced by suitably limiting the reagent. Using alcohols, alkoxy groups are introduced. Either one or all three halogens can be replaced.



Adequate yields in this reaction require extremely efficient removal of HCl; otherwise CH₃Cl is formed. Some commercial processes utilize catalytic reactions of epoxy with POCl₃ (48):



Phosphorus oxychloride is stable to above 300°C. Hydrolysis with water yields phosphoric acid.



Phosphorus oxychloride has strong donor properties toward metal ions. The remarkably stable POCl₃–AlCl₃ complex has been utilized to remove AlCl₃ from Friedel-Crafts reaction products. Any POX₃ molecule contains a pyramidal PX₃ group; the oxygen atom occupies the fourth position to complete the distorted tetrahedron (37). Some properties of phosphorus oxyhalides are presented in Table 8.

Manufacture. Phosphorus oxychloride has been manufactured by oxidizing phosphorus trichloride. When oxygen is bubbled through liquid phosphorus trichloride, complete absorption of pure oxygen is effected in a 1-m depth. When there is good heat exchange, the rate of oxygen absorption remains practically constant until only 3–5 wt % of phosphorus trichloride remains in the oxychloride. The reaction is inhibited by impurities, especially iron and copper, sulfur compounds, or certain impurities from Cl₂ production process. If these impurities are present, the reaction rate between phosphorus trichloride and oxygen exhibits an induction period and then increases to a maximum, after which it falls steadily as the reaction proceeds. A small amount of dissolved phosphorus in the phosphorus trichloride does not influence the reaction.

Table 8. Physical Properties of Phosphorus Oxyhalides^a

Parameter	POF ₃	POCl ₃	POBr ₃	POF ₂ Cl	POFCl ₂	POF ₂ Br	POFBr ₂
CAS Registry Number	[13478-20-1]	[10021-87-3]	[7789-59-5]	[13769-75-0]	[13769-76-1]	[14014-18-7]	[14014-19-8]
boiling point, °C	40	107	193	3.1	52.9	30.5	110.1
melting point, °C	68	1.25	56	96.4	80.1	84.4	117.2
apical angle, °	107	103.5	108	106	106		
bond length, pm							
P—O	155	145	141	155	154		
P—F	151			151	151		
P—Cl		202		202	200		
P—Br			206				

^a Ref. 37.

Heating a mixture of anhydrous phosphorus pentoxide [1314-56-3] and phosphorus pentachloride produces phosphorus oxychloride. Use of expensive phosphorus pentachloride is obviated by using a mixture of the trichloride and chlorine with the pentoxide. Thus, a manufacturing method consists of the chlorination reaction of the trichloride with the pentoxide:



Specifications. Typical analyses for technical-grade POCl₃ are as follows (46):

Property	Description or value
appearance	water-white liquid
distillation range, °C	
first drop	106.7
1–96 mL	1.0
dry point	107.9
specific gravity, 15.5°C / 15.5°C	1.685

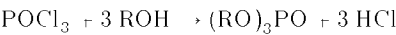
assay, %	99.9
free Cl ₂	trace
PCl ₃ , %	0.10
Fe, ppm	1
crystallizing point, °C	ca 1.1

Storage, Shipping, and Handling. Phosphorus oxychloride is classified by the ICC as a corrosive liquid and a poisonous inhalation hazard. Shipment of POCl₃ must be in conformance with ICC regulations, and individual containers must be affixed with the DOT white acid label and red poison label: DOT UN No. 1810 (46). Phosphorus oxychloride is stored and shipped in 3.8-L (1-gal) or smaller glass containers and DOT-specification wooden overpacking. Bulk POCl₃ shipments are in nickel-clad tank cars of 15,000–30,000 liters (4000–8000 gal) each. Glass and glass-lined steel equipment frequently is used for storage as well as for reaction vessels (47).

Health and Safety Factors, Toxicology. Phosphorus oxychloride volatilizes readily; its vapors are extremely irritating to the eyes, skin, and mucous membranes (49). Direct contact with the liquid can produce severe burns. Contaminated clothing must be removed immediately. Inhalation of POCl₃ vapors can cause pulmonary edema and temporary eyesight problems.

The liquid reacts violently with water, releasing HCl and other gases in sufficient amounts to cause sudden rupture of closed or inadequately vented containers. The acid reaction products can react with metals to generate hydrogen, which is flammable and explosive. The oral LD₅₀ in rats is 380 mg/kg; the inhalation LC₅₀ for rats is 48 ppm/4 h, and for guinea pigs, 53 ppm/4 h (35).

Uses. Phosphorus oxychloride is used extensively to manufacture alkyl and aryl orthophosphate triesters. A slight excess of the respective alcohol or phenol reacts with POCl₃ at elevated temperatures and, if necessary, in the presence of a catalyst.



It is frequently advantageous to favor rapid disengagement of HCl by operating the reaction in an inert solvent such as an alkane. The pure triesters are recovered by a multistep refining process (Fig. 5).

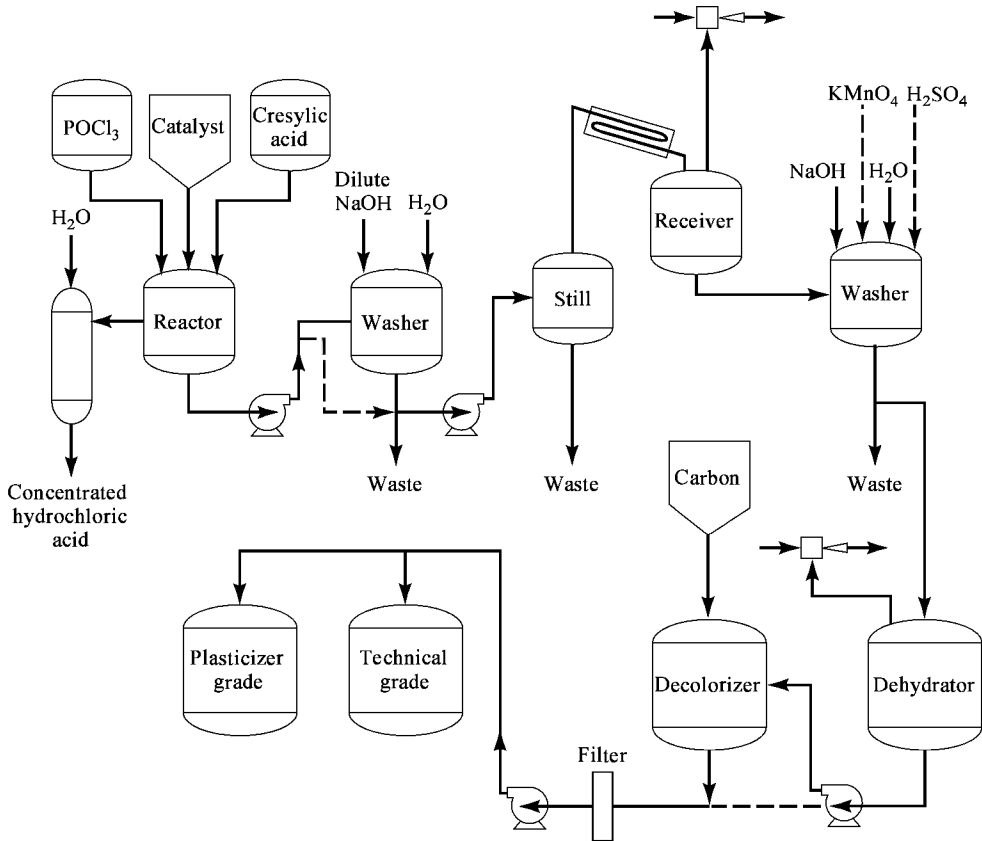


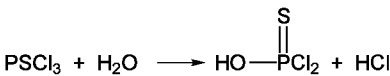
Fig. 5. Manufacturing of triaryl phosphates.

Most of the phosphate esters are used in the production of hydraulic fluids (qv), plastic and elastomer additives, flame retardants (qv), oil stabilizers, pesticides (qv), and medicinal intermediates (see SURFACTANTS). Some trialkyl phosphates, OP(OR)₃, are outstanding solvents for nitrates, especially (UO₂) (NO₃)₂, and therefore are important in uranium processing (see EXTRACTION).

Dialkyl phosphorochloridates, (RO)₂P(=O)Cl, react with trialkyl phosphate esters to give organic pyrophosphates. Organopyrophosphates are anticholinesterase agents and should be handled with great caution (16). Atropine sulfate is a specific antidote.

Phosphorus Sulfochloride.

Properties and Reactions. Phosphorus sulfochloride [3982-91-0] (thiophosphoryl chloride), PSCl₃, is a colorless fuming liquid and is made by the reaction of phosphorus trichloride with sulfur and by the reaction of PCl₅ with P₂S₅. Phosphorus sulfochloride is dimorphic in the solid state. It reacts with water, forming either phosphoric acid or dichlorothiophosphoric acid [14500-94-8], depending on the reaction conditions.



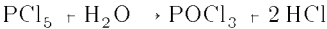
Some physical properties of the phosphorus sulfohalides are summarized in Table 9.

Manufacture. Phosphorus sulfochloride is manufactured by the direct addition of sulfur to phosphorus trichloride (50–52). At about 180°C, the reaction proceeds smoothly. Phosphorus trichloride vapor is passed through an excess of sulfur that is either molten or dissolved in an inert solvent.

The use of alkali or alkaline-earth sulfides catalyzes the reaction so that it is complete in a few hours at 150–160°C; use of aluminum chloride as the catalyst gives a comparable reaction rate at 115°C. When an excess of sulfur is used, the product can be distilled out of the reactor, and the residue of sulfur forms part of the charge in the following batch reaction. The reaction is carried out in a stainless steel autoclave, and the yield is better than 98% based on either reactant. Phosphorus sulfochloride is used primarily in the manufacture of insecticides (53–55), such as Parathion.

Phosphorus Pentachloride.

Properties and Reactions. Phosphorus pentachloride, PCl₅, is a pale, greenish yellow solid having a pungent odor (see Table 7). It is made from PCl₃ and chlorine. Water attacks PCl₅ and the violent hydrolysis proceeds in two stages.



Experimental results suggest that PCl₅ is dimeric in CCl₄ solution. The structure consists of two octahedra sharing edges (56). PCl₅ is monomeric in benzene and apparently is trigonal bipyramidal (36). Solid PCl₅ is ionic, consisting of [PCl₄]⁺ cations and [PCl₆][−] anions (36).

Manufacture. Phosphorus pentachloride is manufactured by either batch or continuous processing. In the former, the phosphorus trichloride usually dissolves in carbon tetrachloride before being treated with chlorine. A mixture of ca one part of phosphorus trichloride to one part of carbon tetrachloride is introduced to a water-jacketed vessel that contains an efficient stirrer and a tight cover with a reflux condenser. The chlorine is passed into the vessel below the liquid level, and crystals of phosphorus pentachloride form in the liquid. When the reaction is completed, the suspension of crystals of phosphorus pentachloride in the carbon tetrachloride is drawn out of the vessel and the crystals are filtered and then dried by circulating hot water through the jacket of the filter. The clarified carbon tetrachloride is returned to the reaction vessel.

Table 9. Physical Properties of Phosphorus Sulfohalides^a

Parameter	PSF ₃	PSCl ₃	PSBr ₃	PSF ₂ Cl	PSFCl ₂	PSF ₂ Br	PSFBr ₂
CAS Registry Number	[2404-52-6]	[3982-91-0]	[3931-89-3]	[2524-02-9]	[2523-93-5]	[13706-09-7]	[13706-10-0]
boiling point, °C	52.9	125	175 (dec)	6.3	64.7	35.5	125.3
melting point, °C	148.8	36.2	39	155.2	96.0	136.9	75.2
apical angles, °	100	101	106			106	100
bond length, pm							
P—S	185	194	189			187	187
P—F	153					145	150
P—Cl		202					
P—Br			213			214	223

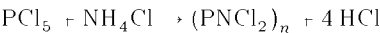
^a Ref. 37.

Specifications and Analysis. Specifications for PCl₅ are as follows: color, light lemon yellow; PCl₃, trace; Ni, 10 ppm max; Pb, 5 ppm max; and Fe, 5 ppm max. Analysis reveals the following: PCl₃, none; Ni, none; Pb, 1 ppm; and Fe, 2 ppm. Bulk density is 0.9 g cm^{−3}.

Storage, Shipping, and Handling. Phosphorus pentachloride is in the EPA extreme hazardous substance list. It is treated as a flammable solid, and containers in which it is stored or shipped must be affixed with a yellow acid label: DOT UN No. 1806. In general, the pentachloride should be handled with the same precautions that are used with the trichloride. Protective clothing should be worn by workers handling the pentachloride and gas masks should be used when personnel are exposed to the vapors.

Health and Safety Factors, Toxicology. Because of its fuming and deliquescent properties, PCl₅ is irritating and corrosive to skin, eyes, and mucous membranes. It reacts with moisture, liberating heat and forming hydrochloric and phosphoric acids which also burn tissue. Inhalation symptoms range from coughing, delayed sneezing, to pulmonary edema. The pentachloride is toxic; its TLV is 1 mg m^{−3} of air. The oral LD₅₀ in rats is 660 mg kg^{−1}; the inhalation LC₅₀ for rats is 205 mg m^{−3}, for mice, 120 ppm 10 min (35). The OSHA standard in air TWA is 1 mg m^{−3}.

Uses.1 Phosphorus pentachloride is used in the manufacture of chlorophosphazenes, and serves as a catalyst and a chlorinating agent in organic syntheses.



Phosphorus Oxides

There are five well-defined oxides of phosphorus: phosphorus(III) oxide [12440-00-5/ P₄O₆, (6); phosphorus(V) oxide [1314-56-3/ (phosphorus pentoxide), P₄O₁₀, (7); phosphorus tetroxide [12164-97-5], P₂O₄; tetraphosphorus heptoxide [12065-80-4], P₄O₇; and tetraphosphorus nonaoxide [12037-11-5], P₄O₉. A rare and higher oxide is the highly oxidizing deep violet solid P₂O₈, which decolorizes and loses oxygen rapidly when heated to 130°C. The structures of P₄O₆ and P₄O₁₀ are related to that of the phosphorus molecule, P₄, (8):

All phosphorus oxides are obtained by direct oxidation of phosphorus, but only phosphorus(V) oxide is produced commercially. This is in part because of the stability of phosphorus pentoxide and the tendency for the intermediate oxidation states to undergo disproportionation to mixtures. Besides the oxides mentioned above, other lower oxides of phosphorus can be formed but which are poorly understood. These are commonly termed lower oxides of phosphorus (LOOPs) and are mixtures of usually water-insoluble, yellow-to-orange, and poorly characterized polymers (58). LOOPs are often formed as a disproportionation by-product in a number of reactions, eg, in combustion of phosphorus with an inadequate air supply, in hydrolysis of a phosphorus trihalide with less than a stoichiometric amount of water, and in various reactions of phosphorus halides or phosphonic acid. LOOPs appear to have a backbone of phosphorus atoms having −OH, =O, and −H pendent groups and is often represented by an approximate formula, (P₄OH)_{*x*}. LOOPs may either hydrolyze slowly, be pyrophoric, or pyrolyze rapidly and yield diphosphine-contaminated phosphine. LOOP can also decompose explosively in the presence of moisture and air near ~150°C.

Phosphorus(V) Oxide.

Properties and Structure. Phosphorus(V) oxide, the extremely hygroscopic acid anhydride of the phosphoric acids, exists in several forms but is often referred to by its empirical formula, P₂O₅. Three crystalline polymorphs, two distinct liquids, and several amorphous or glassy solids are recognized. Some properties of the various forms of phosphoric oxide are listed in Table 10.

Table 10. Properties of Allotropic Forms of Phosphorus(V) Oxide

Allotrope	Melting point, °C	Triple point, °C kPa ^a	Specific gravity	Δ <i>H</i> _{vap} , kJ mol ^b	Structure
-----------	-------------------	-----------------------------------	------------------	---	-----------

				P_4O_{10}	
<i>H</i> (hexagonal)	420	420	480	2.30	95 ^c
<i>O</i> (orthorhombic)	550–570	562	58.3	2.72	152.4
<i>O'</i> (orthorhombic)		580	74.0	2.89	141.9
fused hexagonal (metastable)					67.8
liquid (stable above 580°C)					78.3

^a To convert kPa to psi, multiply by 0.145.
^b To convert J to cal, divide by 4.184.
^c Standard heat of formation is 3009.9 kJ mol^b P₄O₁₀.

The best-characterized form of phosphorus pentoxide is the volatile, metastable, and crystalline *H* (hexagonal) or α -modification. It is obtained in tiny hexagonal flakes when phosphorus pentoxide vapor is slowly cooled. The *H* form has been shown by x-ray diffraction to consist of individual P₄O₁₀ molecules. Although the *H* form is stable indefinitely at room temperature, it changes to the metastable *O* (orthorhombic) form on heating to 400°C in a closed system. The rate of transition is temperature-dependent. Usually, the orthorhombic modification is a chalky aggregate of small crystals. Single crystals are hard and brittle. The crystals of this modification are of infinite sheets in which each phosphorus atom has a tetrahedral set of bonds to three shared and one unshared oxygen atom.

The stable orthorhombic modification *O'* (*T* or δ -form) may be prepared by heating either of the crystalline forms at 450°C for 24 h, producing a horny aggregate of large crystals. Fusion of the material is slow, and superheating of the crystals occurs easily. The melt is a viscous fluid and is different from the melt obtained from the hexagonal form. Studies using x-ray have shown that the *O'* form consists of layers of corrugated sheets of interlocked PO₄ tetrahedra; therefore, it is an infinite polymer of P₄O₁₀, as is the other orthorhombic modification. A high pressure crystalline form is also known. A phase diagram is shown in Figure 6.

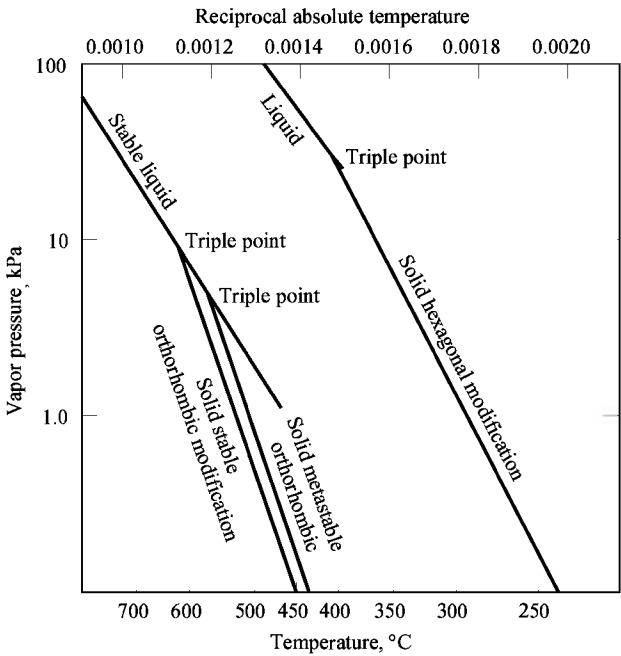


Fig. 6. Phase diagram for three crystalline and two liquid forms of phosphorus(V) oxide. To convert kPa to mm Hg, multiply by 7.5.

Electron diffraction studies indicate that phosphorus pentoxide vapor consists of P₄O₁₀ molecules. The vapor usually condenses to the hexagonal crystalline modification but under rapid cooling can be condensed to an amorphous solid (β -form). The liquid obtained by melting the stable orthorhombic modification cools to form a glass which is the β -form. The liquid obtained from the *H* modification also can be supercooled to a glass.

Manufacture. Phosphorus(V) oxide is made by burning elemental phosphorus in a controlled excess of dry air in a stainless steel, externally cooled combustion chamber similar to that for producing thermal phosphoric acid. The combustion gases are then cooled in a large chamber or barn in which the P₄O₁₀ gas is condensed. Condensation is promoted by rapid mixing with cooled gases. The barn is externally cooled by air or water. The production rate is dependent to a large measure on the ability to cool the barn. The phosphorus pentoxide consists of the hexagonal crystalline modification containing some of the amorphous and *O* forms. If the chamber is water-cooled, the material is a finely divided white powder. However, if the condensation is allowed to take place in a chamber maintained externally at 170–200°C, a denser and more crystalline product is obtained. About 95% of the phosphorus pentoxide is recovered in the barn, the remainder recovered in a phosphoric acid production unit. A fluidized bed for condensing the vapor can be operated to give a variety of phosphorus pentoxide properties (59).

Both technical- and reagent-grade phosphorus pentoxide is typically >99% P₄O₁₀. Phosphorus pentoxide sublimates near 360°C at atmospheric pressure. Lower oxides, which may account for <0.3% (as P₄O) in technical-grade material, are present at <0.02% in reagent-grade phosphorus pentoxide. Lower oxides are detected by decolorization of a dilute potassium permanganate solution (Table 11).

Table 11. Specifications of Technical- and Reagent-Grade Phosphorus(V) Oxide

Parameter	Technical-grade	Reagent-grade ^a
phosphorus pentoxide, wt %	>98.0	>98.0
phosphorus trioxide, wt %	<0.3	<0.02
arsenic, ppm	<50	
iron, ppm	<2	
heavy metals, wt % Pb		<0.01
insoluble matter, wt %		<0.02

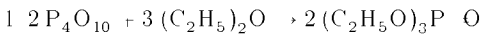
ammonium, wt %	<0.01
color	no reddish tinge
bulk density, g cm ³	~0.7–1.5

^a Ref. 60.

Health and Safety; Storage and Handling. Phosphorus(V) oxide is extremely hygroscopic. It reacts with explosive violence when in contact with water or aqueous solutions. Phosphorus(V) oxide is a local corrosive and irritates skin, eyes, and mucous membranes because of its strong dehydrating action and exothermic formation of phosphoric acid. For example, 10 mg m³ can cause coughing; 1 mg m³ has been suggested as the threshold limit (61). If phosphorus(V) oxide is spilled on body surfaces, it should be brushed off while dry and the skin should be washed immediately with copious amounts of water. Phosphorus(V) oxide is sold in small glass bottles contained in boxes or hermetically sealed metal cans. Larger quantities are shipped in metal barrels or drums. In addition to the yellow label, the Chemical Manufacturers Association (CMA) suggests that a special label warning against burns be affixed to the container.

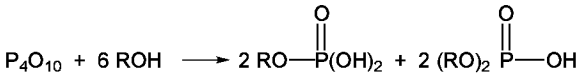
Uses. The most important chemical property of phosphorus pentoxide is its avidity for water. Below 100°C, it is one of the most effective dehydrating agents known. A partial water vapor pressure of the hydrated pentoxide is <10^{−4} Pa at ambient temperature. It reacts with water to form a mixture of condensed phosphoric acids, the composition of which depends on the quantity of water. Phosphorus pentoxide extracts the elements of water from many other substances that are considered good dehydrating agents. For example, phosphorus pentoxide converts pure HNO₃ to N₂O₅ and H₂SO₄ to SO₃, dehydrates amides to nitriles, reacts with alcohols to produce esters of simple and polymeric phosphoric acids, and effects ring closures and other condensation reactions. The main industrial application of phosphorus pentoxide as a drying agent had once been in the dehydration step in the manufacture of methyl methacrylate. It is used as a drying agent for liquids and gases with which it does not react, especially for removing traces of water from vacuum systems (see VACUUM TECHNOLOGY). A drawback in its use as a drying agent is its tendency to coat with a layer of polyphosphoric acid, which prevents further moisture absorption.

Phosphorus(V) oxide is used in the manufacture of phosphorus oxychloride, as a catalyst in air-blowing of asphalt (qv), and as an intermediate for phosphate esters. However, phosphate esters are commonly manufactured from POCl₃, which in turn may be produced from either PCl₃ or P₂O₅ and PCl₅. Phosphate esters are used as surfactants, hydraulic fluids, and plasticizers. Organophosphate surfactants are used widely in industrial and specialty applications, eg, as components of dry-cleaning compounds, water- and oil-based cutting fluids, emulsifiers and wetting agents in textile manufacturing, emulsifiers in emulsion-polymerization processes, pigment dispersants in oil-based paints, emulsifiers and moisture barrier compounds in cosmetics, and mold-release agents. Liquid phosphate esters, eg, tricresyl phosphate [1330-78-5], are used as high temperature stable, fire-resistant hydraulic fluids. A large use is in the preparation of the plasticizer triethyl phosphate [78-40-0].



Phosphorus compounds are effective flame retardants for oxygenated synthetic polymers such as polyurethanes and polyesters. Aryl phosphates and chloroalkyl phosphates are commonly used, although other compounds such as phosphonates are also effective. The phosphorus compounds can promote char formation, thereby inhibiting further ignition and providing an efficient thermal insulation to the underlying polymer.

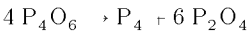
Mixed mono- and dialkyl are used as catalysts for resin curing and as intermediates for fire retardants, oil additives, antistatic agents (qv), and extraction solvents. An equimolar mixture of mono- and dialkyl acid phosphates are formed at a 1:6 mole ratio of oxide to alcohol.



The aryl phosphate esters are similarly produced from phenols. For instance, a mixture of C₉H₁₀–C H₄–(OCH₂CH₂)_{*n*}–O–(OH)₂ and [C₉H₁₀–C H₄–(OCH₂CH₂)_{*n*}–O]₂P(=O)–OH are formed from polyethoxylated nonylphenol. The magnesium salts of the ester mixture are soluble in organics and are used as dry-cleaning surfactants.

Phosphorus(III) Oxide. Phosphorus(III) oxide [12440-00-5], the anhydride of phosphonic acid, is formed along with by-products such as phosphorus pentoxide and red phosphorus when phosphorus is burned with less than stoichiometric amounts of oxygen (62). Phosphorus(III) oxide is a poisonous, white, wax-like, crystalline material, which has a melting point of 23.8°C and a boiling point of 175.3°C. When added to hot water, phosphorus(III) oxide reacts violently and forms phosphine, phosphoric acid, and red phosphorus. Even in cold water, disproportionation may be observed if the oxide is not well agitated, resulting in the formation of phosphoric acid and yellow or orange poorly defined polymeric lower oxides of phosphorus (LOOP).

Phosphorus(III) oxide is slowly oxidized in the air, but when heated above 70°C, it can spontaneously ignite as a result of disproportionation to elemental phosphorus. Above 210°C, the oxide decomposes into phosphorus and phosphorus tetroxide:

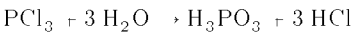


In both the liquid and vapor states, phosphorus(III) oxide exists as the P₄O molecule. The vapor of phosphorus(III) oxide ionizes air. Phosphorus tetroxide is made by heating P₄O in a sealed tube to 440°C. P₄O sublimes under vacuum at 180°C and forms colorless, glossy crystals.



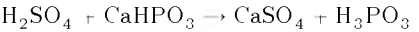
Phosphonic Acid and P(III) Derivatives. Phosphonic or phosphorous acid is a white deliquescent crystalline compound having a melting point of 73.6°C. As evidenced by its structure, HP(=O)(OH)₂, phosphonic acid is dibasic. The first hydrogen is strongly ionized, p*K*_{a1} ca 1.3–1.7, and H₃PO₃ is therefore a stronger acid than orthophosphoric acid. The second dissociation constant, p*K*_{a2}, is ca 6.7. The third hydrogen does not ionize in aqueous solution. Heating phosphonic acid to >ca 250°C results in an exothermic disproportionation to phosphoric acid, phosphine, and hydrogen. A common accident occurs in overheating reactions in which phosphonic acid or LOOP may be formed, eg, in making alkyl chlorides from PCl₃ at too high a temperature. Phosphonic acid is also referred to as phosphorous acid, even though technically the latter is an isomer of phosphinic acid which is found not as the free acid but usually as the triester derivative. The standard heat of formation for phosphonic acid, H₃PO₃(c), is −971.5 kJ mol^{−1} (−232.2 kcal mol^{−1}) (39).

Phosphonic acid is prepared by the dissolution of phosphorus(III) oxide or by the hydrolysis of phosphorus trichloride:



This reaction can be violent partly because of the heat liberated in the solvation of the hydrogen chloride. The hydrolysis can be moderated by adding PCl₃ to a saturated solution of HCl. Subsequently, the water and hydrogen chloride are boiled until the temperature reaches 180°C. On cooling, phosphonic acid crystallizes from the melt.

Additional phosphonic acid is derived from by-product streams. In the manufacture of acid chlorides from carboxylic acids and PCl₃, phosphonic acid or pyrophosphonic acid is produced, frequently with copious quantities of yellow polymeric LOOP. Such mixtures slowly evolve phosphine, particularly on heating, and formerly were a disposal problem. However, purification of this crude mixture affords commercial phosphonic acid. By-product acid is also derived from the precipitate of calcium salts in the manufacture of phosphinic acid. As a consequence of the treatments of the salt with sulfuric acid, carbonate is liberated as CO₂ and phosphonic acid goes into solution.



The acid is recovered from the filtrate after calcium sulfate is removed.

Typical specifications of technical-grade phosphonic acid are given in Table 12. Because disposal of the by-product hydrochloric acid poses problems, several attempts have been made to produce phosphonic acid by a nonhalide route; however, so far none of these efforts have been translated into an industrial process.

Table 12. Specifications for Phosphinic and Phosphonic Acid

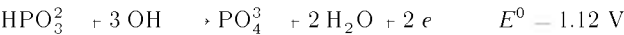
Parameter	Phosphinic acid ^a	Phosphonic acid
assay, %	30–32	69–71
arsenic, ppm ^b	1.5	1
heavy metals, ppm ^b	20	300
iron, ppm ^b		250
barium	pass test	
oxalate	pass test	
chloride, % ^c		<0.01
color, APHA		<40

^a Ref. 63.

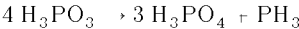
^b Value given is maximum value.

^c As HCl.

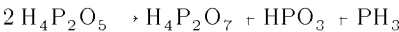
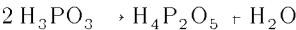
Phosphonic acid and hydrogen phosphonates are used as strong but slow-acting reducing agents. They cause precipitation of heavy metals from solutions of their salts and reduce sulfur dioxide to sulfur, and iodine to iodide in neutral or alkaline solution.



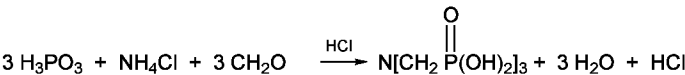
In acid solution, the potential of the H₃PO₃–H₃PO₄ couple is ca +0.20 V. When heated to 200°C, phosphonic acid disproportionates.



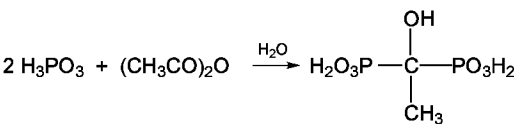
The dimer of phosphonic acid, diphosphonic acid [36465-90-4] (pyrophosphorus acid), H₄P₂O₅, is formed by the reaction of phosphorus trichloride and phosphonic acid in the ratio of 1:5. It is also formed by the thermal decomposition of phosphonic acid. Unlike the chemistry of phosphoric acid, thermal dehydration does not lead to polymers beyond the dimer; extended dehydration leads to a disproportionation to condensed forms of phosphoric acid, such as H₄P₂O₇ [2466-09-3], and phosphine.



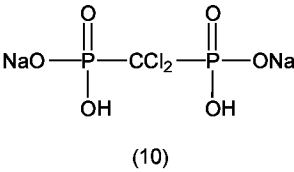
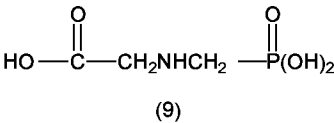
Phosphonic acid is an intermediate in the production of alkylphosphonates that are used as herbicides and as water treatment chemicals for sequestration, scale inhibition, deflocculation, and ion-control agents in oil wells, cooling tower waters, and boiler feed waters. For example, aqueous phosphonic acid reacts with formaldehyde and ammonium chloride in the presence of hydrochloric acid to yield aminotri(methylenephosphonic acid) [6419-19-8].



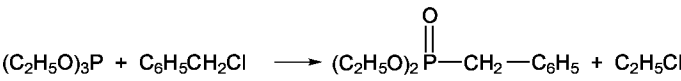
1-Hydroxyethane-1,1-diphosphonic acid[2809-21-4] (HEDP) is produced by hydrolysis of the reaction product of phosphonic acid and acetic anhydride:



Glyphosate, ie, *N*-carboxymethylaminomethanephosphonic acid [1071-83-6] (*N*-phosphonomethyl glycine) (9) is a large-volume, biodegradable, total herbicide sold as the isopropylammonium salt by Monsanto under the trade name of Roundup. 1,1-Bisphosphonates are becoming an important class of pharmaceuticals (qv) for inhibiting bone resorption (calcium regulator). A commercial example is disodium clodronate [22560-50-5] (10) (64).

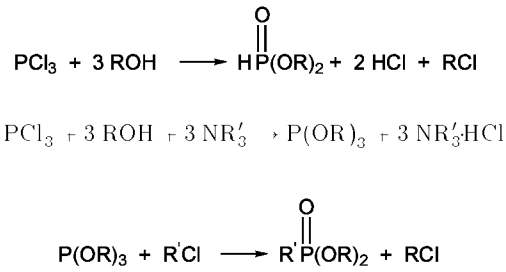


Monoesters of the phosphonic acids are little used in industry. The diesters, O=PR(OR)₂, of phosphonic acid are commonly prepared in industry from trialkyl phosphites in a Michaelis-Arbusov reaction:



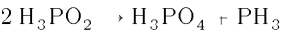
Trialkyl esters of phosphonic acid exist in two structurally isomeric forms. The trialkylphosphites, P(OR)₃, are isomers of the more stable phosphonates, O=PR(OR)₂, and the former may be rearranged to resemble the latter with catalytic quantities of alkylating agent. The dialkyl alkylphosphonates are used as flame retardants, plasticizers, and intermediates. The Michaelis-Arbusov reaction may be used for a variety of compound types, including mono- and diphosphites having aryl as well as alkyl substituents (22). Triaryl phosphites do not readily undergo the Michaelis-Arbusov reaction, although there are a few special cases.

Phosphorus trichloride may also be used directly in the production of trialkyl phosphites, dialkyl phosphonates, and dialkyl alkylphosphonates:

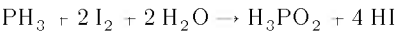


The phosphite triesters are also useful in esterifying carboxylic acids. The triesters are readily oxidized to the respective phosphates.

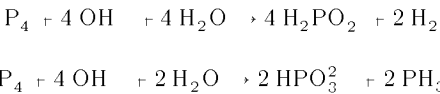
Phosphinic Acid. Phosphinic acid (hypophosphorus acid) is a deliquescent crystalline solid that melts at 26.5°C. It is a monobasic acid having a p*K*_a of 2.1 and the metal salts of which generally exhibit a high solubility. Phosphinic acid disproportionates upon heating above 133°C to generate phosphoric and phosphonic acids, hydrogen, and phosphine.



The standard heat of formation for crystalline H₃PO₂ is −608.8 kJ/mol (−145.5 kcal/mol) (39).
The acid can be prepared by the oxidation of phosphine by iodine and water.



The reaction proceeds quantitatively and the hydroiodic acid can be removed by repeated distillation at 5.3 kPa (40 mm Hg), leaving pure H₃PO₂ as the product. Phosphinic acid may also be prepared by the treatment of barium hypophosphite [14871-79-5] with a stoichiometric quantity of sulfuric acid to precipitate barium sulfate.
Commercially, phosphinic acid and its salts are manufactured by treatment of white phosphorus with a boiling slurry of lime. The desired product, calcium phosphinite [7789-79-9], remains in solution and insoluble calcium phosphite [21056-98-4] is precipitated. Hydrogen and phosphine are also formed, the latter containing sufficient diphosphine to make it spontaneously flammable. The details of this complicated reaction, however, are imperfectly understood. Under some conditions, equal amounts of phosphorus appear as phosphine and phosphite, and the volume of the hydrogen liberated is nearly proportional to the hypophosphite that forms.

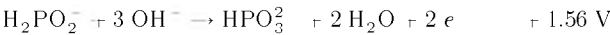
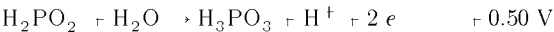


In addition, a small amount of decomposition of hypophosphite by alkali occurs.



Excess calcium hydroxide is precipitated by using carbon dioxide and the calcium carbonate, calcium hydroxide, and calcium phosphite are removed by filtration. The filtered solution is treated with an equivalent amount of sodium sulfate or sodium carbonate to precipitate calcium sulfate or carbonate. Sodium hypophosphite monohydrate [10039-56-2] is recovered upon concentration of the solution. Phosphinic acid is produced from the sodium salt by ion exchange (qv). The acid is sold as a 50 wt % , 30–32 wt % , or 10 wt % solution. The 30–32 wt % solution is sold as USP grade (Table 12) (63).

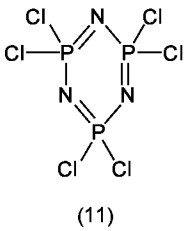
Phosphinic acid and its salts are strong reducing agents, especially in alkaline solution (65).

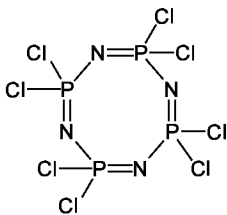


A principal commercial application of the hypophosphites is in the electroless plating (qv) process. Nickel salts are chemically reduced by hypophosphites to form a smooth adherent nickel plating to protect the interiors of large vessels and tank cars. The coating, which can be hardened by heat treatment, usually contains 8–10 wt % phosphorus and is highly impervious.

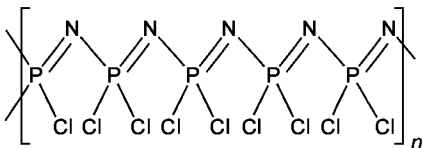
Phosphazenes and Other Phosphorus–Nitrogen Compounds

Phosphazenes. Phosphazenes, (NPX₂)_n, constitute a class of linear and cyclic compounds having an unsaturated skeleton of alternating phosphorus and nitrogen atoms (66,67). Both pure and mixed halophosphazenes and many of their substituted derivatives have been characterized. Most of the species are toxic and irritating. The lower linear chlorophosphazenes are oily substances, whereas most of the cyclic members are white crystalline solids. Hexahalotriphosphazenes, such as hexachlorocyclotriphosphazene [940-71-6], (NPCL₂)₃, (**11**), have almost planar structure with nearly D_{3h} symmetry. The cyclic tetramers, eg, octachlorocyclotetraphosphazene [2950-45-0], (NPCL₂)₄, (**12**), are puckered rings except for the fluoride that is nearly planar. Heating the lower cyclic halophosphazenes results in the formation of linear high polymers such as poly(dichlorophosphazene) [25034-79-1] (**13**). These polyphosphazenes are also known as inorganic rubber because of their elastomeric properties.





(12)



(13)

Hexachlorocyclotriphosphazene and octachlorocyclotetraphosphazene are made by the slow addition of phosphorus pentachloride to an excess of ammonium chloride in a solvent such as chlorobenzene or tetrachloroethane. Addition of 1–10 wt % anhydrous metal chlorides reduces the reaction time to 5–10 h in chlorobenzene or 2–4 h in 1,1,2,2-tetrachloroethane. Suitable chlorides include SnCl₄, MgCl₂, ZnCl₂, MnCl₂, and TiCl₄. Quinoline also acts as a catalyst. The most serious problem is the isolation and purification of the products. Crude mixtures of chlorophosphazenes can be produced quite inexpensively, but the pure cyclic trimer is expensive. The lower cyclic species can be extracted from the reaction mixture by using low boiling petroleum ether; the trimer then is separated by vacuum distillation. Some properties of chlorophosphazenes are listed in Table 13.

Table 13. Physical Properties of Phosphazenes^a

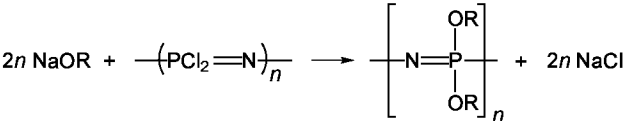
Compound formula	CAS Registry Number	Melting point, °C	Boiling point, °C
Cyclic phosphazenes			
(NPCL ₂) ₃	[940-71-6]	114 ^b	256
(NPCL ₂) ₄	[2950-45-0]	123.5 ^b	328.5
(NPCL ₂) ₅	[14596-41-3]	41.3 ^c	223–224
(NPCL ₂) ₆	[2851-52-7]	92.3 ^c	281–282
(NPCL ₂) ₇	[13827-30-0]	8–18 ^c	289–294
(NPCL ₂) ₈	[14514-98-8]	57–58 ^c	
Linear phosphazenes			
Cl ₃ P=NPCL ₂ N–POCL ₂	[36778-94-6]	32–33	
[Cl(PCL ₂ =N) ₃ PCL ₂] ⁺ (PCL ₂) [–]	[21246-63-9]	95–96	
[Cl(PCL ₂ =N) ₄ PCL ₃] ⁺ (PCL ₂) [–]	[21283-49-8]	98–100	

^a Ref. 66.
^b At 101.3 kPa (1 atm).
^c At 1.7 kPa (13 mm Hg).

The halophosphazenes are hydrolyzed by water. The most reactive species are hydrolyzed on contact with atmospheric moisture at room temperature, but most stable species are hydrolyzed by boiling acid or caustic. The halophosphazenes are degraded to halide ion, phosphoric acid or phosphate, and ammonia or ammonium ion. Although intermediates probably exist between the phosphazene and the decomposition products, they have not been well characterized. The lower chlorophosphazenes are characterized by high vapor pressures at room temperature and small concentrations cause prolonged irritation of eye membranes. Temporary throat and lung irritation following inhalation of these compounds also occurs. Protective clothing is recommended when handling large quantities of chlorophosphazenes.

The cyclic halides can be converted to discrete substitution products by reaction with amines, alcohol, or alkylating agents. For example, (NPCL₂)₃ reacts with ammonia to form (NP(NH₂)₂)₃ [13597-92-7], with *p*-NaOC H₄CH₃ to form (NP(OC H₄CH₃)₂)₃ [27122-73-2], and with CH₃MgCl to form (NP(CH₃)₂)₃ [6607-30-3]. Among the cyclic members, the trimeric halides are the most inert toward substitution and tetrameric halides are the most active.

Poly(dichlorophosphazene), Cl[[bpl]]PCL₂=N[[bpr]]PCL₂[–]X, where X is Cl or PCL₂, is formed by heating purified hexachlorocyclotriphosphazene in the molten state at 210–250°C while protecting the reaction from moisture. The time, temperature, and extent of polymerization to <70% are carefully regulated. An uncross-linked high polymer is formed that is soluble in solvents such as benzene, toluene, or tetrahydrofuran. The chlorine atoms of the dissolved polymer can be efficiently substituted by the addition of nucleophiles to form alkoxy, amino derivatives, etc. For example,



Continued heating of the unsubstituted polymer also yields an insoluble, cross-linked elastomeric polyphosphazenes. The rubber-like material consists of long angled chains with molecular weights on the order of 10⁴–10⁷.

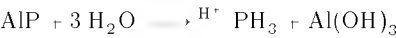
Although polyphosphazenes exhibit interesting properties as elastomers, commercial importance of the unsubstituted chloro polymers has been limited by hydrolytic instability (see ELASTOMERS, SYNTHETIC-PHOSPHAZENES). Polymers substituted with organic side groups, however, are generally stable to water. Properties of the polymers such as crystallinity, hydrophilicity, and electrical conductivity can be controlled over a wide range by selection of the substituted side groups. Properties of the substituted polymers include low temperature flexibility and elasticity; thermal stability in excess of 200°C; high stability to various kinds of radiation; resistance to water, solvents, and oils; nonflammability; and flame retardancy. Hydrocarbon-resistant fluoroalkyloxy elastomers have been fabricated for automotive applications, aryloxy foams were developed for flame resistance as well as heat and sound insulation, and

poly(bis(methoxyethoxyethoxy)phosphazene) [98973-15-0] (MEEP) has been proposed as a solid electrolyte for rechargeable batteries (qv). Biomedical applications such as heart valves and controlled-release drug delivery systems (qv) have also been proposed. However, in spite of their variety of controllable properties, phosphazene polymers have not yet achieved widespread commercial application.

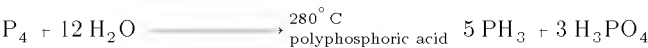
Other Phosphorus–Nitrogen Compounds. Of the binary phosphorus–nitrogen compounds, only triphosphorus pentanitride [12136-91-3], P₃N₅, has been obtained in a pure state. It can be prepared by treatment of P₄S₁₀ with ammonia and subsequent heating of the intermediate product to 850°C in a stream of NH₃. Triphosphorus pentanitride, a solid, is colorless, tasteless, odorless, insoluble in common solvents, stable in ambient atmosphere, and has a specific gravity of 2.51. When heated in vacuum, the compound decomposes above 760°C into the elements, whereas heating in a stream of hydrogen yields P₄ and NH₃. The unstable trichlorophosphineimide [14700-11-9], Cl₃P=NH, results from the reaction of PCl₅ with ammonia. However, if the interaction product of phosphorus pentachloride and ammonia is heated in vacuum, phospham [22722-08-3], (PN₂H)_n, forms. Phospham is a stable, white, and loose powder that is insoluble in water. It appears to have potential utility as a flame retardant, eg, for nylon (68).

Inorganic Phosphines and Phosphides

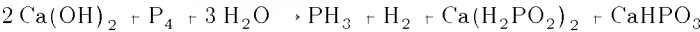
Properties and Reactions. Phosphine [7803-57-2], PH₃, can be produced in a number of ways. However, the inadvertent evolution of phosphine in an otherwise safe reaction is an element of hazard in many procedures involving phosphorus chemicals. Phosphine can be conveniently produced by the hydrolysis of an active metal phosphide, eg, calcium phosphide [1305-99-3], Ca₃P₂, or aluminum phosphide [20859-73-8], AlP (6). Using AlP, yields are high and contamination of diphosphine is minimal.



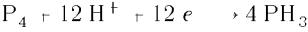
Phosphine is prepared commercially from the acid- or base-catalyzed reaction of elemental phosphorus with water. In the acid-catalyzed reaction, P₄, white phosphorus, converts in part to red phosphorus. The latter is the main reactant (69).



Phosphine is also made as a by-product of the commercial calcium hypophosphite [7789-79-9]. Calcium phosphite [21056-98-4] is also produced.

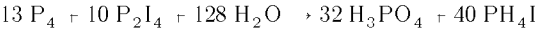


Another approach for the production of phosphine is an aqueous electrolytic process, whereby nascent hydrogen reacts with elemental phosphorus (70). Phosphine is produced at the cathode.



A conductor with a hydrogen overvoltage exceeding that of smooth platinum can be used for the cathode.

Phosphine generated by the above procedures is usually contaminated to varying degrees with diphosphine, which renders it spontaneously flammable. Pure phosphine can be produced by hydrolysis of phosphonium iodide [12125-09-6], PH₄I, which can be made by the action of water on a mixture of phosphorus and diphosphorus tetraiodide [13455-00-0] (71).



The salt, which is separated by sublimation from the other products, can be stored until it is needed and then hydrolyzed.

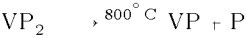


The standard heat of formation of PH₃ (g) is 22.89 kJ mol (5.47 kcal mol) (39).

Phosphides. Compounds of phosphorus containing the more electropositive elements are generally called phosphides. A large number of binary phosphides as well as many ternary mixed-metal phosphides, metal phosphide nitrides, etc, are known. Some binary phosphides, such as those of nickel, exhibit a variety of stoichiometries (Ni₃P, Ni₅P₂, Ni₁₂P₅, Ni₂P, Ni₅P₄, NiP, NiP₂, NiP₃), whereas others, such as aluminum, form only one (AlP). Metalloids such as B and Si also form phosphides.

The phosphides are usually made by direct combination of the elements at elevated temperature. The reactive phosphorus is typically red phosphorus, white phosphorus, or phosphorus vapor. Lithium phosphide [12057-29-3], Li₃P; sodium phosphide [12058-85-4], Na₃P; and potassium phosphide [12260-14-9], KP₁₅, as well as iron(III) phosphide [26508-33-8], FeP, and diiron phosphide [1310-43-6], Fe₂P, are made in this manner.

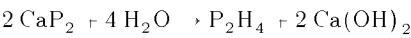
Some phosphides, such as titanium phosphide [12037-65-9], TiP, can be prepared by passing phosphine over the metal or its halide. Reaction of phosphine with heavy metal salt solutions often yields phosphines that may contain unsubstituted hydrogens. Phosphides may also be prepared by reducing phosphorus-containing salts with hydrogen, carbon, etc, at high temperatures, the main example of which is the by-product formation of ferrophosphorus in the electric furnace process for elemental phosphorus. Phosphorus-rich phosphides such as vanadium diphosphide [12037-77-3] may be converted to lower phosphides, eg, vanadium phosphide [12066-53-4], by thermal treatment.



Phosphides have varying degrees of metallic, covalent, and ionic characters to the bonding. Strongly electropositive metals, eg, alkali metals, alkaline-earths, and lanthanides, yield ionic phosphides that react readily with water to generate phosphine. The metal usually remains as hypophosphite or hydroxide. The phosphine generated from hydrolysis of the phosphides is usually contaminated with pyrophoric diphosphine, which renders the evolved gas spontaneously flammable at a high enough concentration.

The usual alkali metal phosphides are reddish brown to black and are stable up to 650°C but react instantly with moisture to form phosphine. Magnesium phosphide [12057-74-8], Mg₃P₂, and aluminum phosphide, AlP, or the mixed compounds, are stable in dry air but decompose on contact with water or humid air. These compounds, prepared by direct union of the elements, are used with an igniting agent, eg, 1 wt % nitric acid or nitric oxide, in sea flares. The evolved phosphine burns, producing a dense cloud of phosphoric acid mist and a highly luminous flame. Aluminum phosphide has been used since the 1930s for the generation of phosphine in grain fumigation (see WHEAT AND OTHER CEREAL GRAINS). The resulting phosphine gas is deadly to pests but does not exhibit long-lasting deleterious effects. Aluminum phosphide can be shipped and stored readily as tablet formulations having a protective coating of paraffin to prevent accidental hydrolysis. Magnesium phosphide, calcium phosphide [1305-99-3], Ca₃P₂, and zinc phosphide [1314-84-7], Zn₃P₂, are also used as fumigants and or rodent poisons. Calcium phosphide is prepared commercially by heating quick lime in phosphorus vapor.

The alkali metal phosphides of formula M₃P and the alkaline-earth phosphides of formula M₃P₂ contain the P³⁻ anion. Calcium diphosphide [81103-86-8], CaP₂, contains P⁴⁻₂, and reaction with water liberates diphosphine and maintains the P–P linkage.



The existence of a variety of other polyphosphide anions has been demonstrated. For example, P⁵⁻₃ and P⁷⁻₅ have been identified in LaP₂ (72). Infinite chain structures for the polyphosphide anion have been reported for compounds such as KP₁₅ and thallium pentaphosphide [11093-99-5], TlP₅. Cage anions such as P³⁻₇ that include triply linked phosphorus atoms are found in Li₃P₇.

Phosphides of the less electropositive metals and the metalloids may be considered more as metal–phosphorus alloys. These are thermally stable

and typically resistant to attack by water, even at 100°C. Some are resistant to dilute acids or bases but most are decomposed by stronger oxidizing acids or by bases. Transition-metal-rich phosphides are comparable in structure and properties to transition-metal borides and silicides.

Phosphides are sometimes classified according to stoichiometry into metal-rich phosphides, ie, $M : P > 1$; monophosphides, $M : P = 1$; and phosphorus-rich phosphides, $M : P < 1$. The metal-rich phosphides are typically hard and brittle, having a metallic luster and processing high electrical and thermal conductivities. The monophosphides are also typically hard, chemically inert, and possess a luster. Many of the monophosphides and the phosphorus-rich phosphides are semiconductors (qv). Gallium phosphide [12063-98-8], GaP, and indium phosphide [22398-80-7], InP, are specialty semiconductors having a high electron–hole mobility. These phosphides and related ternary compounds are used in light-emitting diodes (qv), solar energy (qv) conversion panels, field effect transistors, photodetectors (qv), etc. The structures of the phosphorus-rich phosphides are made of increasingly polymerized polyphosphide anions as the $M : P$ ratio decreases.

Ferrophosphorus is produced as a by-product in the electrothermal manufacture of elemental phosphorus, in which iron is present as an impurity in the phosphate rock raw material. The commercial product contains ca 23–29% P and is composed primarily of Fe_2P [1310-43-6] and Fe_3P [12023-53-9], along with impurities such as Cr and V. Ferrophosphorus is used in metallurgical processes for the addition of phosphorus content. Low concentrations (up to ~0.1%) of phosphorus in wrought and cast iron and steel not only increases the strength, hardness, and wear resistance but also improves the flow properties. In large structural members and plates, it is desirable to use a type of steel that does not need to be quenched or tempered, and thus does not exhibit weld-hardening. This property is afforded by the incorporation of a small quantity of phosphorus in steel. Ferrophosphorus from western U.S. phosphorus production is used as a raw material for the recovery of vanadium (see VANADIUM AND VANADIUM ALLOYS).

Copper and tin phosphides are used as deoxidants in the production of the respective metals, to increase the tensile strength and corrosion resistance in phosphor bronze [12767-50-9], and as components of brazing solders (see SOLDER AND BRAZING ALLOYS). Phosphor bronze is an alloy of copper and 1.25–11 wt % tin. As tin may be completely oxidized in a copper alloy in the form of stannic oxide, 0.03–0.35 wt % phosphorus is added to deoxidize the alloy. Phosphor copper [12643-19-5] is prepared by the addition of phosphorus to molten copper. Phosphor tin [66579-64-4], 2.5–3 wt % P, is made for the deoxidation of bronzes and German silver.

Organophosphines and Derivatives

Preparation and Properties of Organophosphines. Aliphatic phosphines can be gases, volatile liquids, or oils. Aromatic phosphines frequently are crystalline, although many are oils. Some physical properties are listed in Table 14. The most characteristic chemical properties of phosphines include their susceptibility to oxidation and their nucleophilicity. The most common derivatives of the phosphines include halophosphines, phosphine oxides, metal complexes of phosphines, and phosphonium salts. Phosphines are also raw materials in the preparation of P^I derivatives, ie, derivatives of the isomers phosphinic acid, HP(OH)₂, and phosphonous acid, H₂P(=O)OH.

Table 14. Physical Properties of Phosphines^a

Compound	CAS Registry Number	Boiling point, °C	Density at 20°C, g cm ³	Dipole moment, 10 ^{−30} Cm ^b	Index of refraction, n _D ²⁰	pK _a	nmr Parameters	
							³¹ P chemical shift, ppm	J _{P-H} Hz ^c
PH ₃	[7803-51-2]	88	1.529 ^d				238–241	180–182
CH ₃ PH ₂	[593-54-4]	14		3.7			163	188
(CH ₃) ₂ PH	[676-59-5]	21		4.10			100	188–191
(CH ₃) ₃ P	[594-09-2]	38–41			1.192	8.65 ^e	62–63	
C ₂ H ₅ PH ₂	[593-68-0]	25		3.87			128	185
(CH ₃ CH ₂) ₂ PH	[627-49-6]	85–86	0.7862	4.7	1.447	33.7 ^f	56	190
(CH ₃ CH ₂) ₃ P	[554-70-1]	129–130	0.7999	4.7–9.7	1.456	8.8 ^e	20–21	
<i>n</i> -C ₄ H ₉ PH ₂	[1732-74-7]	86–88	0.7693	4.54	1.4372	0.03 ^e	135	195
(<i>n</i> -C ₄ H ₉) ₂ PH	[1732-72-5]	178–186	0.8083		1.456	4.51 ^e	70	180
(<i>n</i> -C ₄ H ₉) ₃ P	[998-40-3]	240–242	0.817–0.82	5.0–7.3	1.4635	8.43 ^e	32	
iso-C ₄ H ₉ PH ₂	[4023-52-3]	60–80				0.02 ^e		
<i>tert</i> -C ₄ H ₉ PH ₂	[2501-94-2]	66–67					120	182
(<i>n</i> -C ₈ H ₁₇) ₃ P	[4731-53-7]	291 ^g			1.4683		32–33	
CF ₃ PH ₂	[420-52-0]	25.5		6.41			129	201
HOCH ₂ CH ₂ PH ₂	[16247-01-1]	139–140	1.004		1.4950			
HO ₂ CCH ₂ PH ₂	[4757-26-0]	85–86 ^h					143	198
H ₂ NCH ₂ CH ₂ PH ₂	[14085-97-3]	110					150	194
C H ₅ PH ₂	[638-21-1]	157–160	1.001 ⁱ	3.70	1.5796	24.5 ^f	119–124	195–201
(C H ₅) ₂ PH	[829-85-6]	280				0.03 ^e	41.4	214–239
(C H ₅) ₃ P	[603-35-0]	384 ^j		4.74–5.14		2.73 ^e	6–8	

^a Refs. 8, 73, and 74.

^b To convert Cm to debye, divide by 3.336 × 10^{−30}.

^c J_{P-H} is the coupling constant between the phosphorus and hydrogen atoms.

^d Value is in g L at 0°C and 101.3 kPa (1 atm).

^e In H₂O.

^f In CH₃OH.

^g At 6.7 kPa (50 mm Hg); melting point = 48°C.

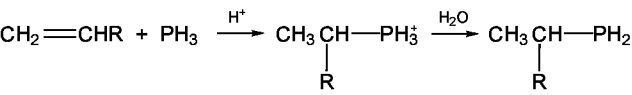
^h At 1.3 kPa (10 mm Hg).

ⁱ At 15°C.

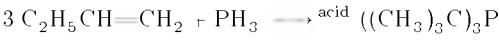
^j Melting point = 79–81°C.

There are a few economical routes that can be employed for production of the largest-volume phosphines as specialty chemicals. The preparation of alkyl phosphines, where R > C₂H₅, employs the addition of lower phosphines across an olefinic double bond. The reaction may be either acid-, base-, or radical-catalyzed. The acid-catalyzed addition probably proceeds through the generation of a carbonium ion intermediate which is attacked by the unshared

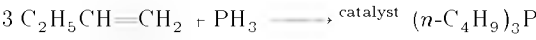
pair on phosphorus. A typical Markovnikoff distribution of products occurs as follows:



The free phosphine is liberated upon the removal of the acid catalyst with water. Tri-*t*-butylphosphine [998-40-3] is prepared by the acid-catalyzed addition of isobutene to phosphine.

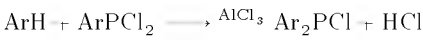
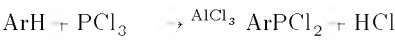


In the presence of strong bases, the reaction is thought to proceed via the formation of a negatively charged phosphide in a Michael mechanism. Free-radical initiators are also effective at reduced temperatures. Free-radical reactions proceed through ready homolysis of the P–H bonds. At >300°C, there appears to be sufficient dissociation of PH₃ to initiate a free-radical addition in the absence of a catalyst. If low molecular weight olefins are present, some carbon-chain growth also may occur. Linear trialkyl phosphines are the usual products in base- or radical-catalyzed additions because they are thermodynamically favored.

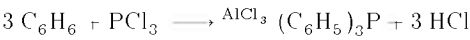


In the presence of a large excess of PH₃, primary phosphines, RPH₂, are formed predominantly. Secondary phosphines, R₂PH, must be either isolated from mixtures with primary and tertiary products or made in special multistep procedures. Certain secondary phosphines can be produced if steric factors preclude conversion to a tertiary product. Both primary and secondary phosphines can be substituted with olefins. After the proper selection of substituents, mixed phosphines of the type RR'PH or RR'R'P can be made.

Aryl (Ar) phosphines are conveniently prepared by Friedel-Crafts reactions.

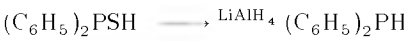
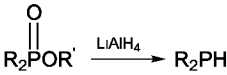
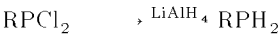


Halophosphines can be reduced to phosphines. Various methods have been employed to recover the product, but yields are only fair. Triphenylphosphine [603-35-0] is prepared from benzene and phosphorus trichloride:

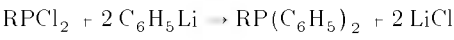
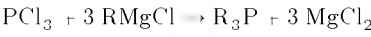


The crude products can be purified by distillation. Highly pure triphenylphosphine can be obtained by recrystallization. Stoichiometric amounts of AlCl₃ are generally required because of the complexation by the phosphorus compound.

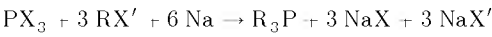
In general, if the desired carbon–phosphorus skeleton is available in an oxidized form, reduction with lithium aluminum hydride is a powerful technique for the production of primary and secondary phosphines. The method is applicable to halophosphines, phosphonic and phosphinic acids as well as their esters, and acid chlorides. Tertiary and secondary phosphine oxides can be reduced to the phosphines.



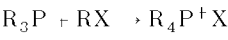
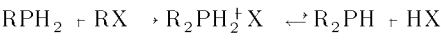
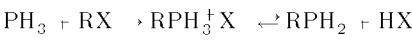
Halophosphines can be treated with organometallic reagents, eg, Grignard reagents, organomercuries, and organolithiums, to produce tertiary phosphines.



Yields are best in the case of aromatic metallic reagents. Use of aliphatic reagents favors low molecular weight products. Products often are recovered by water addition, followed by separation and distillation of the organic layer. Such procedures inevitably lead to acidic by-products when there is incomplete replacement of the halogens on phosphorus. A modification of the Wurtz reaction sometimes is used.

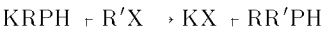
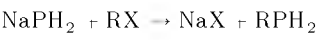


The addition of alkyl halides to phosphines is analogous to the reactions with amines. Because primary phosphonium salts are highly dissociated, the reaction proceeds to the tertiary or quarternary salts.

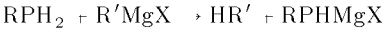


Alkylation of phosphines by alkyl halides exhibits reactivity relative to the base strength, ie, PH₃ is the least reactive and tertiary phosphines the most. This reactivity reflects the difficulty in using alkylation to prepare anything except quaternary phosphonium halides.

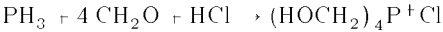
Metal phosphides can be employed to direct the action of alkyl halides more toward primary and secondary phosphines.



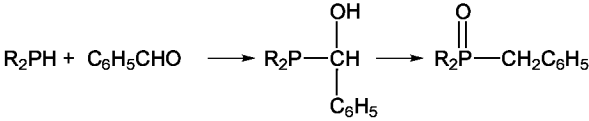
Products are contaminated with more highly alkylated compounds, but less so than without the active metal. The phosphide can be generated from a Grignard or organolithium reagent.



The addition of P–H bonds across a carbonyl function leads to the formation of α -hydroxy-substituted phosphines. The reaction is acid-catalyzed and appears to be quite general with complete reaction of each P–H bond if linear aliphatic aldehydes are used. Steric considerations may limit the product to primary or secondary phosphines. In the case of formaldehyde, the quaternary phosphonium salt [124-64-1] is obtained.



If aromatic aldehydes or ketones are used, the tertiary phosphine product sometimes rearranges to a mixed phosphine oxide.

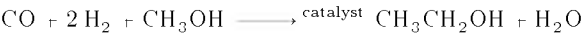


Other methods of preparation have been described in detail (8,75,76).

Aliphatic phosphines can be gases, volatile liquids, and oils. Aromatic phosphines frequently are crystalline, although many are oils. Physical properties of selected phosphines are listed in Table 14.

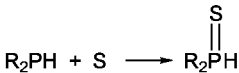
Phosphines readily coordinate with typical Lewis acids, eg, proton, BF₃, and metals form compounds such as PH₄I, the adduct (CH₃)₃PBF₃ [1898-77-7], and the complex [(C H₅)₃P]₂NiC₈H₁₂ [36510-64-2], respectively. Phosphines usually are considered soft bases and interact most strongly with soft acids or highly polarizable electron-pair acceptors (14). Complexation is facilitated by the back-donation of electrons from the acceptor to the unfilled *d* orbitals of phosphorus (77). Because groups on the phosphorus are also polarizable, they can reinforce the availability of the unshared electron pair. Thus, triphenylphosphine and other aromatic phosphines are much better ligands than might be anticipated from the relatively poor coordinating ability of the anilines and other aromatic amines. The base strength of phosphines increases with the degree of substitution. With respect to the proton, the relative base strengths of the methyl-substituted phosphines are PH₃ < CH₃PH₂ < (CH₃)₂PH < (CH₃)₃P. The quaternary phosphonium hydroxide [4814-27-8], ((CH₃)₄P)⁺OH[–], is a strong base and essentially is ionized 100% in water. The p*K*_a values of phosphines in water have been correlated with Taft's Σσ^{*} function, which is a measure of the inductive effect (78). Representative p*K*_a values are listed in Table 14.

Because of phosphines' wide-ranging bonding ability to acceptors such as metals, these are broadly used ligands in the construction of catalysts. A phosphine can block a specific site on a central metal under all conditions, but it can block a site only in some species and not in others. It can also make a metal more soluble and thus dispersable in a liquid medium. Varying steric requirements for a reactive site on a catalyst can be achieved by adjusting the bulk and configuration of the groups on the phosphorus ligand (8). For example, [(*n*-C₄H₉)₃PPtCl₂]₂SnCl₂ is a useful hydroformylation catalyst for α -olefins. This complex minimizes the production of branched-chain products (79). Catalysts derived from cobalt carbonyl and tributylphosphine converts internal olefins to linear alcohols (80). Rhodium complexes of chiral phosphines, eg, methylpropylphenylphosphine [4653-62-7], can be used as stereoselective hydrogenation catalysts (81). The complex [Co(CO)₃(P(C₄H₉)₃)₂]⁺[Co(CO)₃][–] provides a high yield of cyclooctene from cyclooctadiene (82). The presence of a phosphine and a nickel catalyst, which is used to homologize butadiene, produces cyclooctadiene, whereas the same catalyst without the phosphine produces cyclododecatene (83). The use of tertiary alkylphosphines in conjunction with the following reaction reduces the formation of by-products (84).

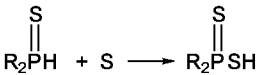


A mixture of (C₄H₉)₃P, TiCl₃, and AlCl₃ is useful for polymerizing C₃–C₄ olefins (85). The dimerization of propylene is accomplished by using catalysts such as Ni(PR₃)₄ (86). Alkylphosphines such as *tert*-butylphosphine [2501-94-2] have been proposed as a substitute for high purity phosphine in the production of the semiconductor gallium phosphide (87).

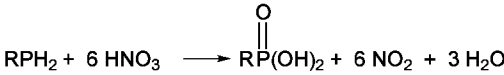
Reactions of Phosphines. Phosphines are generally subject to air oxidation at ambient temperatures via a free-radical mechanism. Because of higher stability, tertiary aryl phosphines are the slowest to air-oxidize but undergo oxidation readily at higher temperatures. Phosphines also react with elemental sulfur and other common oxidizing agents such as H₂O₂ or HNO₃. Secondary or tertiary phosphines yield the corresponding oxides or sulfides, eg, triphenyl phosphine yields triphenyl phosphine oxide [791-28-6].



The secondary phosphine oxides or sulfides can be oxidized to phosphinic acids or thiophosphinic acids.

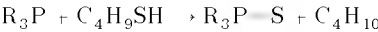
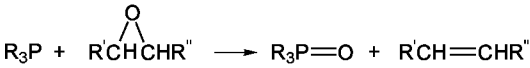


Primary phosphines are oxidized as intermediates in the syntheses of phosphonic acids and phosphonates.



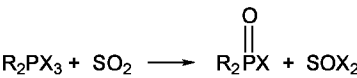
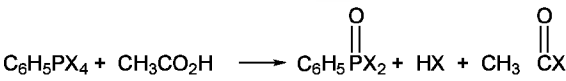
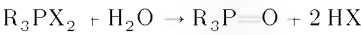
Secondary phosphines are likewise intermediates in the syntheses of phosphonites. The primary and secondary phosphines can also be used to make other mixed phosphines.

Because relatively mild oxidizing agents react with phosphines, the latter are convenient deoxidizers (88) or desulfurizers (89):

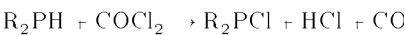


Functional groups within the substituents in a phosphine usually behave similarly to a hydrocarbon, provided that they do not react with the phosphine group.

Pentavalent phosphorus derivatives can be converted to phosphonyl halides or phosphine oxides by partial hydrolysis or by other oxygen donors.

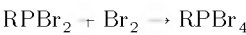
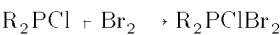
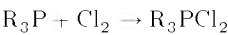


Primary and secondary phosphines can be treated with halogenating agents to produce halophosphines.



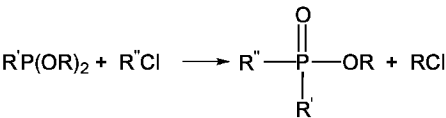
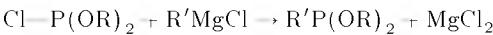
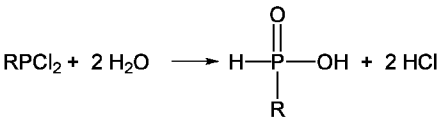
Phosgene, COCl₂, is a useful chlorinating reagent, but PCl₅, SOCl₂, SO₂Cl₂, ClSO₃H, and SiCl₄ also are used. The activity of elemental halogens must be limited by low temperatures, dilute reactions, and carefully controlled additions so that only trivalent derivatives form.

Tertiary phosphines and primary and secondary phosphines can be oxidized by elemental halogen to halophosphine halides.

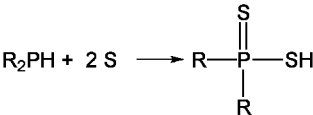


The reactions are general for chlorine and bromine. Iodine does not react to form pentavalent phosphorus compounds. Fluorides are best formed with less active fluorinating agents.

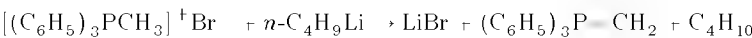
Dihalophosphines or halophosphites, prepared from phosphorus trichloride, are used in the synthesis of organophosphinates.



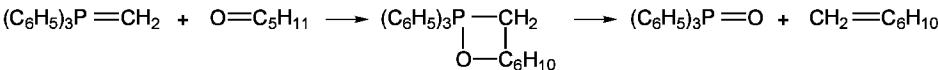
Organophosphinates (90) may also be prepared by the oxidation of secondary phosphines or halophosphines with hydrogen peroxide or sulfur:



A useful application of phosphines for replacing a carbonyl function with a carbon-carbon double bond is the Wittig reaction (91). A tertiary phosphine, usually triphenylphosphine, treated with the appropriate alkyl halide which must include at least one α-hydrogen, yields the quaternary salt [1779-49-3] which is then dehydrohalogenated to form the Wittig reagent, methylenetriphenylphosphorane [19943-09-5], an ylide.



Sodium hydride, sodium amide, or other strong bases also can be used. The reagent can be generated in the presence of an appropriate carbonyl compound that reacts directly.



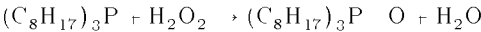
Although the reaction is a general one for aldehydes and ketones, in certain instances, esters, anhydrides, and nitrites are attacked by these reagents (8,92).

Health and Safety Factors, Toxicology. Because low molecular weight phosphines generally are spontaneously flammable, they must be stored and handled in an inert atmosphere. The upper explosion limit is 1.6% and the upper limit is near 100% (93). The higher and less volatile homologues are more slowly oxidized by air and present less of a problem.

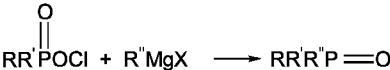
Phosphine is a central nervous system and liver toxin. Subacute poisoning produces blood and lymph stasis; subchronic poisoning is characterized by degeneration of ganglion cells as well as disintegration of the gums. Phosphine is one of the most toxic of the simple gases; it is lethal to adults in a 0.5–1-h exposure at 0.05 mg/L (8). The corresponding value for H₂S is 9.6 mg/L; for HCN, 0.12 mg/L. Federal specifications regarding exposure are 400 mg/m³ TWA and the TLV is 0.3 ppm (35). The LC₅₀ for phosphine is 11 ppm·h for rats. The lowest reported LC₅₀s are 2500 ppm·20 min for rabbits, and 1000 ppm·5 min for humans. For dibutylphenylphosphine [6372-44-7], the lowest reported LC₅₀ is 1100 mg/m³ in 10 min. Phosphine is unacceptable for transport on either passenger or cargo aircraft.

Phosphine Oxides. Controlled oxidation of secondary or tertiary phosphines using H₂O₂ yields the corresponding phosphine oxides. Control

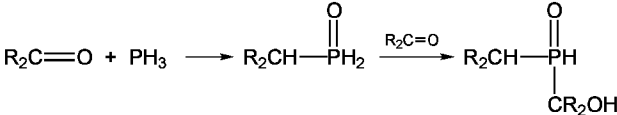
of the reaction temperature must be maintained to prevent the further oxidation to the phosphinic acids. Trioctylphosphine oxide [78-50-2] (TOPO) may be manufactured by the radical-catalyzed addition of 1-octene to trioctyl phosphine [4731-53-7], followed by peroxide oxidation.



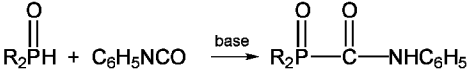
Phosphine oxides having higher alkyl substituents are also prepared industrially using Grignard reagents.



Phosphine oxides may be prepared by the acid-catalyzed reaction of phosphine with carbonyl compounds such as ketones (94).

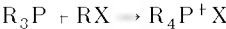


Because of their relative instability, primary phosphine oxides cannot be isolated and must be converted directly to derivatives. Primary and secondary phosphine oxides undergo reactions characteristic of the presence of P–H bonds, eg, the base-catalyzed nucleophilic addition to unsaturated compounds such as olefins, ketones, and isocyanates (95).



Tertiary phosphine oxides are stable. The temperatures required for thermal decomposition are approximately 300°C higher than the corresponding amine oxides (96). Trimethyl phosphine oxide is stable to 700°C.

Phosphonium Salts. The most common route to phosphonium salts is the reaction of tertiary phosphines with alkyl or aryl halides in polar solvents.



Phosphonium salts may also be prepared by the addition of tertiary phosphines to carbonyl compounds or olefins (97).

■

Tetrakis(hydroxymethyl)phosphonium hydroxide, used for flameproofing cellulosic fabrics, is manufactured in a two-step process.



Curing the treated fibers with ammonia chemically attaches the compound to the cloth. The corresponding sulfate has replaced much of the hydroxide because under certain conditions of manufacture or use the carcinogen bis(chloromethyl) ether may form.

Phosphonium salts are typically stable crystalline solids that have high water solubility. Uses include biocides, flame retardants, the phase-transfer catalysts (98). Although their thermal stability is quite high, tertiary phosphines can be obtained from pyrolysis of quaternary phosphonium halides. The hydroxides undergo thermal degradation to phosphine oxides as follows:



Economic Aspects

Phosphorus compounds are manufactured for a variety of uses, either directly or as intermediates in the production of other compounds. Manufacturing of the largest-volume products is summarized in Table 15, and prices are given in Table 16. Phosphorus trichloride and phosphorus pentasulfide are the compounds in highest demand. Phosphorus trichloride production increased steadily from 1985 through the mid-1990s. Up to 36% of PCl₃ is used for pesticide products.

Table 15. U.S. Production of Phosphorus Compounds, t × 10³^a

Compound	1980	1985	1988	1990	1992	1994
PCl ₃	87	77	104	128	127	143
P ₂ S ₅	73	66	65	63	59	55
POCl ₃	31	25	27	28	25	31
P ₂ O ₅	6.4	6.0	5.6	5.1	5.1	5.4

^a Ref. 99.

Table 16. Prices of Phosphorus Compounds, \$/kg^a

Year	PCl ₃	P ₂ S ₅	POCl ₃	P ₂ O ₅
1970	0.23	0.29	0.25	0.38
1975	0.73	0.66	0.82	1.23

1980	0.97	0.88	1.06	1.73
1986	0.37	1.06	0.88	1.87
1987	0.71	1.15	0.95	1.81
1988	0.71	1.15	0.95	1.80
1989	0.84	1.23	0.95	1.80
1990	0.90	1.35	1.10	1.85
1991	0.95	1.35	1.23	1.85
1992	0.95	1.50	1.22	1.85
1993	1.01	1.50	1.23	1.85
1994	1.01	1.42	1.30	1.85

^a Ref. 100.

Twenty-five percent of PCl_3 is used for the manufacture of surfactants and sequestrants, which are used widely in industrial and specialty applications, eg, heavy-duty liquid detergents, components of dry-cleaning compounds, water- and oil-based cutting fluids, emulsifiers and wetting agents in textile manufacturing, emulsifiers in emulsion–polymerization processes, pigment dispersants in oil-based paints, emulsifiers and moisture barrier compounds in cosmetics (qv), and mold-release agents (see DISPERSANTS; PAINT). In Europe, phosphonate surfactants have been used in some household detergent formulations (101).

Organophosphorus compounds, primarily phosphonic acids, are used as sequestrants, scale inhibitors, deflocculants, or ion-control agents in oil wells, cooling-tower waters, and boiler-feed waters. Organophosphates are also used as plasticizers and flame retardants in plastics and elastomers, which accounted for 22° of PCl_3 consumed. Phosphites, in conjunction with liquid mixed metals, such as calcium–zinc and barium–cadmium heat stabilizers, function as antioxidants and stabilizer adjuncts. In 1992, such phosphorus-based chemicals amounted to slightly more than 6° of all such plastic additives and represented 8500 t of phosphorus. Because PVC production is expected to increase, the use of phosphorus additive should increase 3° annually through 1999.

Phosphonic (phosphorous) acid, produced by hydrolysis of PCl_3 , is for the most part consumed captively. It has also been offered as a flaked product and a 70 wt ° solution by Rh  ne-Poulenc. Phosphonic acid is a by-product from manufacturing carboxylic acid chlorides and alkali peroxides. Additional by-product phosphonic acid is recovered in the manufacture of phosphinic acid.

Phosphorus trichloride is also used in the manufacture of antifoam agents, catalysts, dyes and pigments, as well as pharmaceutical and quaternary compounds, and is commonly used as a chlorinating agent. Phosphorus trichloride is used to make phosphorus oxychloride, which is used in the manufacture of adsorbents for air filters, antifoam agents, dyes and pigments, mineral-processing materials, pharmaceuticals (qv), and solvents. These uses represented 32,000 t of PCl_3 in 1988 and 30,000 t in 1994.

Liquid phosphate esters, eg, tricresyl phosphate [1330-78-5], are one of two types of fire-resistant hydraulic fluids (qv). Fire-resistant fluids account for less than 10° of the total fluids market. Phosphorus-based fluids generally are stable at high temperatures in addition to being fire resistant. Approximately 10,000 t of organophosphorus compounds were used in hydraulic fluids in 1994. The manufacture of these materials consumed ca 4000 t of POCl_3 .

Approximately 35° or 220,000 t of the total U.S. 1992 pesticide production were phosphorus-containing products. These were based on the primary products, PCl_3 and P_2S_5 , and on the secondary products, POCl_3 , PSCl_3 , and H_3PO_3 . The pesticide market accounted for about 27,000 t (43°) of the U.S. P_2S_5 production in 1993. This is a decline from about 36,000 t in 1980. Organophosphate insecticides offer broad spectrum pest control. For this reason, organophosphate insecticides, eg, Malathion, Parathion, etc, are under environmental pressure. Further, organophosphate insecticides face strong competition from pest control agents having higher specificity and much lower application rates, such as synthetic pyrethroids. Synthetic pyrethroids are chemical equivalents of natural pyrethrins, which are toxins produced in plants and insects that provide defense against (other) insects (see INSECT CONTROL TECHNOLOGY; PESTICIDES). As a result, the use of P_2S_5 for insecticides is expected to be flat or declining in the United States.

Detergents, dispersants, antiwear agents, extreme pressure agents, corrosion inhibitors, oxidation inhibitors, and viscosity-index improvers are all classified as lubricating oil additives (LOAs). The lubricating oil additive market accounted for about 36,000 t (57°) of the U.S. P_2S_5 production in 1993, an increase from ca 32,000 t in 1980. The bulk of organophosphate oil additives are zinc and other metal dialkyl and diaryl phosphorodithioates, $\text{Zn}[(\text{RO})_2\text{PS}_2]_2$. These compounds provide antiwear, anticorrosion, and antioxidant properties to the oil. No other compound has been identified that performs all three functions. Further, there is no known substitute for crankcase antiwear. However, there is some concern for the effect of organophosphate oil additives on catalytic converters in automobiles. The use of P_2S_5 for oil additives is expected to grow with the economy.

Although over 1.1×10^6 t of phosphorus pentoxide was produced in 1992, 99.5° of this material was converted immediately into phosphoric acid. The amount of phosphorus pentoxide produced and sold as such in the United States in 1994 was 6200 tons. The typical 1994 price was \$1.6–1.7 kg. The only U.S. producer is Rh  ne-Poulenc at Mt. Pleasant, Tennessee. Phosphorus pentoxide, which is used to make asphalt-blowing agents and in water treatment, amounts to 1.0×10^3 t in 1992.

Approximately 4500 tons of sodium hypophosphite [7681-53-0], NaH_2PO_2 , was produced in 1990. This material is used principally in electroless nickel plating of plastic objects. Of the secondary products made from primary phosphorus compounds, phosphorus oxychloride is manufactured in the largest volume. Phosphorus pentachloride and phosphorus sulfochloride are made from phosphorus trichloride.

BIBLIOGRAPHY

"Phosphorus Compounds, Inorganic" and "Phosphorus Compounds, Organic" in *ECT* 1st ed., Vol. 10, pp. 461–510, J. R. Van Wazer, Monsanto Co.; "Phosphorus Compounds" in *ECT* 2nd ed., Vol. 15, pp. 298–320, by J. R. Van Wazer, Monsanto Co.; in *ECT* 3rd ed., Vol. 17, pp. 490–539, by I. A. Boenig, M. M. Crutchfield, and C. W. Heitsch, Monsanto Co.

1. *Proceedings of the XII International Conference on Phosphorus Compounds*, Vols. 75–77, Toulouse, France, 1992.
2. *Naming and Indexing of Chemical Substances for Chemical Abstracts, Chemical Abstracts Service*, American Chemical Society, Washington, D.C., 1992.
3. *Nomenclature of Organic Chemistry*, IUPAC, Pergamon Press, New York, 1979.
4. A. J. de Bethune and N. A. S. Loud, *Standard Aqueous Electrode Potentials and Temperature Coefficients at 25 °C*, C. A. Hampel, Skokie, Ill. 1964.
5. L. E. Sutton, ed., *Tables of Interatomic Distances and Configuration in Molecules and Ions*, The Chemical Society, London, U.K. 1958.
6. J. R. Van Wazer, *Phosphorus and Its Compounds*, Vol. 1, Interscience Publishers, New York, 1958.
7. G. M. Kosolapoff, *Organophosphorus Compounds*, John Wiley & Sons, Inc., New York, 1950; G. M. Kosolapoff and L. Maier, *Organic Phosphorus Compounds*, Wiley-Interscience, New York, 1972.
8. R. F. Hudson, *Structure and Mechanism in Organophosphorus Chemistry*, Academic Press, Inc., New York, 1965; S. Trippett, ed., *Organophosphorus Chemistry: A Review of the Literature Published*, 10 Vols., The Chemical Society, London, U.K., 1970 to 1980; *Chimie Organique du Phosphore*, Colloques Internationaux du Centre National de la Recherche Scientifique 182, Paris, France, May, 1969.
9. M. Grayson and E. J. Griffith, eds., *Topics in Phosphorus Chemistry*, 11 Vols., Wiley-Interscience, New York, 1964 to 1983.
10. D. G. Gorenstein and F. H. Westheimer, *Proc. Nat. Acad. Sci. U.S.* **58**, 1747 (1967); R. S. Berry and co-workers, *Acta Chem. Scand.* **22**, 231 (1968).
11. T. Imamoto, in R. Engel, ed., *Handbook of Organophosphorus Chemistry*, Marcel Dekker, Inc., New York, 1992, pp. 1–53; W. J. Stec, in S. Oae, A.

- Ohno, and N. Furukawa, eds., *Reviews on Heteroatom Chemistry*, Myu, Tokyo, 1988, pp. 367–389.
12. E. Niecke and D. Gudat, *Angew. Chem., Intl. Ed. Engl.* **30**, 217 (1991); M. Regitz, *Chem. Rev.* **90**, 191 (1991).
13. D. E. C. Corbridge, *Phosphorus: An Outline of Its Chemistry, Biochemistry, and Technology*, 4th ed., Elsevier, Science, Inc., New York, 1990, p. 43.
14. F. Basolo and R. G. Pearson, *Mechanisms of Inorganic Reactions*, 2nd ed., John Wiley & Sons, Inc., New York, 1967.
15. G. Kamai, V. A. Kukhtin, and O. A. Strogova, *Trudy Kazan. Khim. Tekhnol. Inst. im. S. M. Kirova*, (21), 155 (1956).
16. R. D. O'Brien, *Toxic Phosphorus Esters*, Academic Press, Inc., New York, 1960.
17. A. A. Durrant, T. G. Pearson, and P. L. Robinson, *J. Chem. Soc.* 730 (1930); **Brit. Pat. 990,918** (Mar. 29, 1963), E. J. Lowe and F. A. Ridgeway (to Albright and Wilson Mfg., Ltd.).
18. S. E. Frazier, R. P. Nielsen, and H. H. Sisler, *Inorg. Chem.* **3**, 292 (1964).
19. G. M. Kosolapoff, *Organophosphorus Compounds*, John Wiley & Sons, New York, 1950, p. 16.
20. **Ger. Offen. 2,061,838** (June 19, 1972), H. Geffers and co-workers (to Bayer); Fr. Demande **2,118,698** Sep. 1, 1972 (to Bayer).
21. **Ger. Pat. 3,829,961** (Mar. 8, 1990), R. Kleinstueck and co-workers (to Bayer).
22. A. Michaelis and R. Kaehne, *Chem. Ber.* **31**, 1048 (1898); A. E. Arbuzov, *J. Russ. Phys. Chem. Soc.* **38**, 687 (1906); G. M. Kosolapoff, *J. Am. Chem. Soc.* **67**, 1180 (1945).
23. K. Moedritzer and R. R. Irani, *J. Org. Chem.* **31**, 1603 (1966); U.S. Pat. 3,257,479 (June 21, 1966), R. Irani and K. Moedritzer (to Monsanto Co.).
24. D. A. Bright, F. Jaffe, and E. N. Walsh, in E. N. Walshemps and co-workers, eds., *Phosphorus Chemistry Developments in American Science*, American Chemical Society, Washington, D.C., 1992.
25. H. R. Allcock and T. J. Fuller, *J. Am. Chem. Soc.* **103**, 2250 (1981); H. R. Allcock, P. J. Harris, and R. A. Nissan, *J. Am. Chem. Soc.* **103**, 2256 (1981).
26. H. Hoffman and M. Becke-Goehring, in E. Griffith and M. Grayson, eds., *Topics in Phosphorus Chemistry*, Vol. 8, Wiley-Interscience, New York, 1976, p. 193.
27. **U.S. Pat. 2,794,705** (June 4, 1957), R. B. Hudson (to Monsanto Co.).
28. **Eur. Pat. 439,391 A1** (July 31, 1991), P. Engel & A. Courant (to Atochem).
29. **U.S. Pat. 3,183,062** (May 11, 1965), J. A. Taylor (American Agricultural Co.).
30. J. Cremer, *Chem. Ing. Tech.* **37**, 705 (1965).
31. R. Schumann, A. Zschalich, and H. Matschiner, *Chemische Technik* **34**(6), 294 (1984).
32. **U.S. Pat. 4,664,896** (May 12, 1987), R. E. Hall (to FMC).
33. M. C. Demarcq, *Phosphorus Sulfur*, **11**, 65 (1981).
34. M. C. Demarcq, *Ind. Eng. Chem. Res.* **30**, 1906 (1991).
35. D. V. Sweet, ed., *Registry of Toxic Effects of Chemical Substances-1988*, U.S. Dept. of Health and Human Services, Center for Disease Control, National Institute for Occupational Safety and Health, Washington, D.C., 1987.
36. F. A. Cotton and G. Wilkinson, *Advanced Inorganic Chemistry*, Wiley-Interscience, New York, 1988.
37. P. J. Durrant and B. Durrant, *Introduction to Advanced Inorganic Chemistry*, 2nd ed., John Wiley & Sons, Inc., New York, 1970.
38. R. Remy, *Lehrbuch der Anorganischen Chemie*, 13th ed., Akademische Verlagsgesellschaft, Leipzig, Germany, 1970.
39. D. D. Wagman and co-workers, *The NBS Tables of Chemical Thermodynamic Properties; Selected Values for Inorganic and C₁ and C₂ Organic Substances in SI Units*, in *J. Phys. Chem. Ref. Data*, **11**, suppl. 2 (1982); M. W. Chase, Jr. and co-workers, *JANAF Thermochemical Tables*, 3rd ed., Part II, in *J. Phys. Chem. Ref. Data*, **14**, suppl. 1, 1985.
40. F. Hossenlopp, M. McPartlin, and J.-P. Ebel, *Bull. Soc. Chim. France*, 2219 (1965).
41. **U.S. Pat. 3,361,528** (Jan. 2, 1968), C.-Y. Shen (to Monsanto Co.).
42. C.-H. Yang, "Froth Phenomena in a Phosphorus Trichloride Reactor," American Institute of Chemical Engineers annual meeting, Miami Beach, Fla., 1992.
43. **U.S. Pat. 5,260,026** (Nov. 9, 1993), S. F. Macauloy and co-workers (to Monsanto Co.); **U.S. Pat. 5,310,529** (May 10, 1994), S. F. Macauloy and co-workers (to Monsanto Co.).
44. **U.S. Pat. 5,252,306** (Oct. 12, 1993), J. L. Champion, C.-Y. Shen, and C.-H. Yang (to Monsanto Co.).
45. **U.S. Pat. 5,208,000** (May 4, 1993), C.-H. Yang, and D. O. Fisher (to Monsanto Co.).
46. *Phosphorus Trichloride and Phosphorus Oxychloride*, Monsanto Chemical Intermediates Co., St. Louis, Mo., 1977.
47. *PCl₃/POCl₃ Handling Guidelines*, Monsanto Co., St. Louis, Mo., 1990.
48. A. D. F. Toy and E. N. Walsh, *Phosphorus Chemistry in Everyday Living*, 2nd ed., American Chemical Society, Washington, D.C., 1987.
49. *Properties and Essential Information for Safe Handling and Use of Phosphorus Oxychloride*, Chemical Manufacturers Association Chemical Safety Data Sheet SD-26, Washington, D.C., 1948.
50. **Ger. Pat. 218757** (Feb. 13, 1985), P. Reichelt and co-workers, (to VEB Chemiekombinat Bitterfeld).
51. Jpn. Pat., 53,116,295 (Oct. 11, 1978), Y. Matsushita and co-workers (to Sumitomo Chemical).
52. **Ger. Pat. 2,032,832** (Feb. 18, 1971), H. Weinelt, G. Henning, and K. J. Becker (to VEB Chemiekombinat Bitterfeld).
53. P. C. Vyas and co-workers, *Asian J. Chem.* **5**, 844 (1993).
54. **Eur. Pat. 437,335** (July 17, 1991), R. M. Jacobson, L. T. Nguyen, and J. R. Ramsey (to Rohm & Haas).
55. R. L. Gupta and K. Nripendra, *Pestic. Sci.* **15**, 553 (1984).
56. V. P. Petro, *Cryoscopic, Spectroscopic, and Conductimetric Study of Phosphorus Pentachloride in Selected Non-Aqueous Solvent Systems*, University Microfilms, Ann Arbor, Mich.
57. *Phosphorus Pentachloride*, Twin Lake Chemical Inc., Lockport, N.Y., 1995.
58. P. Royen and K. Hill, *Z. Anorg. Chem.* **229**, 97 (1936).
59. **U.S. Pat. 4,777,025** (Oct. 11, 1988), W. Kowalski and co-workers (to Hoechst).
60. *ACS Specifications, Reagent Chemicals*, American Chemical Society, Washington, D. C., 1993.
61. *Properties and Essential Information for Safe Handling and Use of Phosphoric Anhydride*, Chemical Manufacturers' Association Chemical Safety Data Sheet SD-28, Washington, D.C., 1974.
62. **U.S. Pat. 4,980,142** (Dec. 25, 1990), J. D. McGilvery, H. Twardowska, and S. M. Cybulski (to Tenneco Canada).
63. *The United States Pharmacopeia* 23 (USP 23-NF18), The United States Pharmacopeial Convention, Inc., Rockville, Md., 1994.
64. E. Pohjala and co-workers, *Phosphorus Sulfur Silicon*, **76**, 159 (1993); *The Merck Index*, 11th ed., Merck and Co., Rahway, N.J., 1989.
65. W. M. Latimer, *Oxidation Potentials*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1952.
66. H. R. Allcock, *Phosphorus-Nitrogen Compounds*, Academic Press, Inc., New York, 1972; J. E. Mark, H. R. Allcock, R. West, *Inorganic Polymers*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1992.
67. M. Zeldin, K. J. Wynne, and H. R. Allcock, eds., *ACS Symp. Ser.* **360**, 250–302 (1988); P. Wisian-Neilson, K. J. Wynne, and H. R. Allcock, eds., *ACS Symp. Ser.* **572**, 208–342 (1988).
68. E. D. Weil and N. Patel, *Fire Materials*, **18**, 1 (1994).
69. **Brit. Pat. 990,918** (Mar. 19, 1963), E. J. Lowe and F. A. Ridgeway (to Albright and Wilson Mfg., Ltd.).
70. **U.S. Pat. 3,109,795** (Nov. 5, 1963), I. Gordon (to Hooker Chemical Corp.); **U.S. Pat. 3,109,787** (Nov. 5, 1963), D. T. Price and I. Gordon (to Hooker Chemical Corp.); **U.S. Pat. 3,109,788** and 3,109,789 (Nov. 5, 1963), G. T. Miller and J. Steingart (to Hooker Chemical Corp.).
71. J. B. Work, *Inorg. Syn.* **2**, 141 (1946).
72. H. G. Von Schneering, M. Wittmann, and D. Sommer, *ZAAC*, **510**, 61 (1984).

73. V. Mark and co-workers, in E. J. Griffith and M. Grayson, eds., *Topics in Phosphorus Chemistry*, Wiley-Interscience, New York, 1967, p. 290.

74. D. R. Lide, ed., *Handbook of Chemistry and Physics*, 72nd ed., The Chemical Rubber Company, Boca Raton, Fla., 1991.

75. S. Trippett, ed., *Organophosphorus Chemistry: A Review of the Literature Published*, The Chemical Society, London, U.K., 1980.

76. *Chimie Organique du Phosphore, Colloques Internationaux du Centre National de la Recherche Scientifique* 182, Paris, May 1969.

77. A. B. Berg and R. I. Wagner, *J. Am. Chem. Soc.* **75**, 3872 (1953).

78. W. A. Henderson, Jr., and C. A. Strueli, *J. Am. Chem. Soc.* **82**, 5791 (1960).

79. **U.S. Pat. 4,101,565** (July 18, 1978), J. E. Poist (to Celanese Corp.).

80. **U.S. Pat. 3,907,909** (Sept. 23, 1975), A. Macaluso and O. W. Ridgeway (to Texaco Oil Co.).

81. **U.S. Pat. 3,968,147** (July 6, 1976), A. J. Solodar (to Monsanto Co.); **U.S. Pat. 3,849,480** (Nov. 19, 1974), W. S. Knowles and M. J. Sabacky (to Monsanto Co.).

82. **U.S. Pat. 3,717,585** (Apr. 3, 1970), D. R. Fahey (to Phillips Petroleum Corp.).

83. G. Wilke and co-workers, *Adv. Chem. Ser.* **34**, 137 (1962); **Fr. Pat. 1,320,729** (Mar. 15, 1963), G. Wilke and co-workers (to Studiengesellschaft Kohle mbH).

84. **Ger. Offen. 2,625,627** (Dec. 30, 1976), L. H. Slaugh (to Shell International Research Corp.); **U.S. Pat. 4,133,966** (Jan. 9, 1979), W. R. Pretzer, P. T. Kobylinski, and J. E. Bronzils (to Gulf Oil Co.).

85. **U.S. Pat. 3,992,320** (Dec. 7, 1976), P. Schneider, H. Schick, and H. Mueller-Tamm (to BASF).

86. **U.S. Pat. 4,115,470** (Sept. 19, 1978), D. R. Fahey and J. E. Mahan (to Phillips Petroleum Corp.).

87. S. H. Li and co-workers, *J. Cryst. Growth*, **96**, 906 (1989).

88. M. J. Baskin and D. B. Denney, *Chem. Ind. (London)*, 330 (1959).

89. M. J. Baskin and D. B. Denney, *J. Amer. Chem. Soc.* **82**, 4736 (1960).

90. **Brit. Pat. 2,068,381** (Aug. 12, 1981), A. J. Robertson and T. Ozog (to Cyanamid Canada).

91. G. Wittig and U. Schoilkopf, *Chem. Ber.* **87**, 1318 (1954); E. Vedejs, *Science*, **207**, 42 (Jan. 4, 1980).

92. H. O. House, *Modern Synthetic Reactions*, W. A. Benjamin, Inc., Menlo Park, Calif., 1972.

93. H. Ohtani and co-workers, *Combust. Flame*, **76**, 307 (1989).

94. S. A. Buckler and M. Epstein, *J. Amer. Chem. Soc.* **82**, 2076 (1960).

95. M. M. Rauhut and H. A. Currier, *J. Org. Chem.* **26**, 4626 (1961).

96. W. J. Bailey, W. M. Muir, and F. Marktscheffel, *J. Org. Chem.* **27**, 4404 (1962).

97. H. Hoffmann, *Chem. Ber.* **94**, 1331 (1961).

98. H. H. Freedman, *Pure Appl. Chem.* **58**, 857 (1986).

99. M. Smart, M. Jaeckel, and Y. Ishikawa, in *Chemical Economics Handbook Marketing Research Report*, SRI International, Menlo Park, Calif., 1995.

100. *Phosphorus Trichloride*, in *Chemical Products Synopsis*, Mannsville Chemical Products Corp., Asbury Park, N.J., June, 1992; *Phosphorus Oxychloride*, *Ibid.*, July, 1992; *Phosphorus Pentasulfide*, *Phosphorus Oxide*, *Ibid.*, Jan., 1989.

101. H. B. May, H. Nijs, and V. Godecharles, "Phosphonates Multifunctional Ingredients for Laundry Detergents," 59th Annual Convention of the Soap and Detergent Association, Boca Raton, Fla., 1986.

General References

E. N. Walsh and co-workers, *Phosphorus Chemistry Developments in American Science*, American Chemical Society, Washington, D.C., 1992.

R. Engel, ed., *Handbook of Organophosphorus Chemistry*, Marcel Dekker, Inc., New York, 1992.

G. Bettermann and co-workers, "Phosphorus Compounds, Inorganic," and J. Svara, N. Weferling, and T. Hofmann, "Phosphorus Compounds, Organic," in B. Elvers, S. Hawkins, and G. Schulz, eds., *Ullmann's Encyclopedia of Industrial Chemistry*, Vol. A19, VCH Verlagsgesellschaft mbH, Weinheim, Germany, 1991.

D. E. C. Corbridge, *Phosphorus: An Outline of Its Chemistry, Biochemistry, and Technology*, 4th ed., Elsevier, Science, Inc., New York, 1990.

F. R. Hartley, ed., *The Chemistry of Organophosphorus Compounds*, Vols. 1–4, John Wiley & Sons, New York, 1990.

R. Engel, *Synthesis of Carbon–Phosphorus Bonds*, CRC Press, Boca Raton, Fla., 1988.

R. S. Edmundson, ed., *Dictionary of Organophosphorus Compounds*, Chapman and Hall, New York, 1988.

A. D. F. Toy and E. N. Walsh, *Phosphorus Chemistry in Everyday Living*, 2nd ed., American Chemical Society, Washington, D.C., 1987.

R. S. Edmundson, "Phosphoric Acid Derivatives," *Compr. Org. Chem.* **2**, 1257 (1979).

G. M. Kosolapoff and L. Maier, *Organic Phosphorus Compounds*, 7 vols, Wiley-Interscience, New York, 1972 to 1976.

A. A. Eldridge, G. M. Dyson, A. J. E. Welch, and D. A. Pantony, eds., *Mellor's Comprehensive Treatise on Inorganic and Theoretical Chemistry*, Vol. VIII, Suppl. III, Longman, London, U.K., 1971.

Ref. 6; J. R. Van Wazer, *Phosphorus and Its Compounds*, Vol. 2, Interscience Publishers, Inc., New York, 1961.

G. M. Kosolapoff, *Organophosphorus Compounds*, John Wiley & Sons, New York, 1950.

T. J. Farr, *Phosphorus-Properties of the Element and Some of Its Compounds*, The Tennessee Valley Authority, Wilson Dam, Ala., 1950.

Organophosphorus Chemistry, Vols. 1–25, Specialist Periodical Reports, The Royal Society of Chemistry, Cambridge, U.K., 1969–1994.

Darrell C. Fee
David R. Gard
Chen-Hsyong Yang
Monsanto Company

PHOTOCHEMICAL TECHNOLOGY

Survey,
Photocatalysts,

SURVEY

Intense research in photochemistry has provided a substantial base of photochemical technologies for industrial application. The state-of-the-art in applied photochemistry has been reviewed through 1989 (1) and subsequently benchmarked at the International Symposium on Perspectives in Photochemistry in 1992 (2). The oldest recorded continuously practiced application of photochemical technology is the processing of natural dyes of the indigo class as used in the traditional Japanese textile craft of aizome (see DYES,NATURAL; TEXTILES) (3). Practice of aizome, popular for over 10 centuries, is no longer of commercial importance.

Photoimaging is another use of photochemical technology (see IMAGING TECHNOLOGY). The first known photographic image was recorded by Niépce in 1826 by the photocross-linking of bitumen (see COAL) coated in a thin layer on a glass plate; exposure of the plate took several days in a camera oscura (4). Bitumen is a naturally occurring mixture of oligomerized, aromatic, and unsaturated hydrocarbons which undergo cross-linking on absorption of the ultraviolet component of sunlight. Oil of lavender was used as a developer to remove uncross-linked hydrocarbon from unexposed regions of the plate to reveal the negative image (see PHOTOGRAPHY). This photochemical innovation gave rise to modern photopolymer technology which is responsible for the environmentally significant solventless coatings (qv) industry (5,6).

Photochemical technology has been developed so as to increasingly exploit inorganic and organometallic photochemistries (2,7), recognizing the importance of photoinduced electron transfer as the phenomenological basis of a majority of commercially successful photochemical technologies (5,8). Use of coherent light sources in industrial applications has led to the field of photodynamic therapy as a photochemically based medical technology (9–11). The application of photochemistry to information storage and communication processes is expected (12) (see INFORMATION STORAGE MATERIALS; RESIST MATERIALS).

Light Sources

Conventional, incoherent light sources suitable for industrial-scale photochemistry and the reactors exploiting them have been reviewed in depth (2). Subsequent improvements in traditional light sources have been incremental.

The exploitation of coherent sources has been a much awaited advance in photochemical technology. Application of laser sources has proved revolutionary in photoimaging (6), and enabled innovations in photodynamic therapy (9–11) and photochemical memory technologies (12). As of this writing (ca 1995), lasers (qv) have not yet proved to be of significant importance for large-scale industrial synthetic applications, but have been applied to effect various photochemistries, eg, photodecomposition of polymers, surface treatments, high precision machining of synthetic polymeric structures, and medical surgery (13). The latter two applications are examples of laser ablation, the use of laser energy to decompose solid structures to gaseous products. The mechanism of ablation has been established to involve primarily photochemical reactions (14). Availability of high powered CO₂ lasers for ablation applications on the one hand, and of electronically modulatable solid-state lasers for high bandwidth information storage, imaging, and communications applications on the other hand, both emitting in the red and near infrared (0.6–9 μm), has stimulated renewed interest in infrared photochemistry.

In 1988 the excimer lamp, a new kind of lamp, was announced (15). An excimer lamp is a gas-phase fluorescent lamp powered by an electrical (15) or microwave (16–18) discharge. It produces monochromatic, incoherent radiation in a quasi-continuous mode, ie, pulses of light at a relatively high (ca 100 Hz) repetition rate. Wavelength of emission is the same as for the corresponding excimer laser, eg, 248 nm for KrF (16), 193 nm for ArF (17), and 157 nm for F₂ (18). Efficiencies of these lamps are theoretically as high as 40% and under practical operating conditions efficiencies of ca 10% based on microwave power are obtained (16–18) for the more efficient, microwave discharge powered lamps. One application for excimer lamp photochemistry is the hardening and or drying on-press of uv-curable inks used in the printing industry, eg, for newspaper printing (19) (see RADIATION CURING). Such inks (qv) are capable of the appropriate rheological characteristics for web offset printing at lower (or zero) levels of volatile solvents and, after uv-curing do not smear in contact with press components, offset to other printed surfaces, or come off on the readers' hands. Excimer lamps are also used to cure epoxy adhesives (qv) incorporating, eg, diaryliodonium or analogous trarylsulfonium salts as photo-acid sources, in the process of making laminated products such as industrial wipers, work wear, surgical drapes, etc (20). Principal application of excimer lamp technology, however, has been in the photochemical deposition of insulating layers in microelectronic fabrication.

Photophysics

Photochemical Laws and Sensitization. All photochemical technologies are practiced in accord with the fundamental laws of photophysics, relating to the nature of light and its interaction with matter. These are covered in detail in texts and reference books on photochemistry (see *General References*). Of interest herein are the following: (1) the Bunsen-Roscoe law, ie, only light which is absorbed by the reactive system is useful; (2) Einstein's law, ie, light absorption is quantized and each quantum absorbed activates one molecule; and (3) the lowest energy excited state of a given spin multiplicity is the starting point for practical photochemical processes. A corollary of the first principle is that not only must the desired component of a photochemical reaction mixture absorb light from the source to be exploited but also other components of the mixture must not absorb in the same spectral regime.

A special case applies when the photochemical reaction mixture is in the form of a thin film, eg, a radiation curable coating, a microlithographic resist, an imaging medium, or a thin, flowing film of a solution in a flow reactor. Accordingly the active light absorbing component in the mixture should optimally be present at a concentration such as to absorb ca 63% of the incident actinic radiation. Inefficient utilization of the radiation results at lower concentration and reaction occurs nonuniformly at higher concentrations (21).

In many cases of practical interest, the reactive species does not absorb light from available sources. In these cases a sensitizer is used that is capable of absorbing available radiation and subsequently transferring either the excitation energy or, in many cases, an electron to the reactive system. In principle the sensitizer does not participate in the reaction and should not affect its course. In practice this is often not the case. Diaryliodonium salts are commonly used photoinitiators for either free-radical or cationic polymerizations (5,6,22) (see INITIATORS). The diaryliodonium cations typically do not absorb significantly at wavelengths longer than 280 nm. Rather, trifluoromethylacetophenone, which absorbs the Hg arc line at 313 nm, sensitizes iodonium salt photolysis by triplet state energy transfer, ie, the excitation energy of the sensitizer is transferred to the iodonium cation. In this case, the sensitizer is recovered unchanged. Electron-transfer sensitization is often more useful in many industrial applications than energy-transfer sensitization. Thus polycyclic aromatic hydrocarbons, eg, anthracene which absorbs out to 400 nm, overlapping the emission of readily available relatively inexpensive medium pressure Hg arc lamps, also sensitize diaryliodonium cations, but sensitization occurs by electron transfer. Because efficient in-cage recombination consumes all the radical products of the reaction, anthracene sensitization is useful only for photoinitiation of cationic cures, eg, of epoxies, via the protic acid by-product.

Previous expositions of photochemical laws have distinguished prominently between states of singlet and triplet multiplicity (1). This distinction continues to be important with respect to photophysics of small organic molecules, but among inorganic and organometallic compounds, states of other multiplicities, eg, doublet and quartet states (23), play an important role. Spin conservation characterizes electronic molecular excitations and localized

excited state species. From transition-state theory a pseudo-equilibrium constant, K^* , is defined:

$$K^* = \frac{h\nu}{k_b T}$$

where h is Planck's constant, k_b is the Boltzman constant, and T is the temperature in Kelvin. By convention, k is specified in units of $L^2 \text{ mol}^{-2} \text{ s}^{-1}$. Then a free energy of activation can be defined as follows:

$$\Delta G^\ddagger = k_b T \ln K^*$$

There are two approaches to estimation of $\Delta G^\ddagger$. The first is an empirical approach (36) based on dynamics of fluorescence quenching of aromatic hydrocarbons in acetonitrile solution. Accordingly,

$$\Delta G^\ddagger = \left[(\Delta G^\circ)^2 + (\lambda/4)^2 \right]^{1/2} + G^\circ/2$$

where the adjustable parameter, $\lambda = 0.42 \text{ eV}$ (40.4 kJ mol^{-1}), for all reactive pairs studied. Subsequent investigations have shown that λ is solvent dependent. This equation is often called the Rehm-Weller equation.

In work for which Marcus and Levich separately received the Nobel Prize, it was shown from first principles (37) and in different contexts that

$$\Delta G^\ddagger = (\lambda/4) (1 + \Delta G^\circ/\lambda)^2 + W$$

wherein λ was defined as a medium reorganization energy which is related to the energy thermalized during relaxation of the excited state to the *thexi*-state (28). The value of λ is predictable in many cases where medium relaxation is the principal contributor to λ by a dielectric continuum approximation. The variable W incorporates the work of bringing the reactants together and separating the products.

This expression, usually called the Marcus equation, is generally applicable to all one-electron transfer processes whether photoinduced or not. Dynamics of electron transfer reflect the driving force provided by photon energy absorption in the ΔG term according to the Weller equation. Thus, if the parameters of the Weller equation can be established experimentally, only estimation of λ is required for prediction of k . It can be seen on inspection that the empirical Rehm-Weller expression is a special case of the Marcus expression; this relationship and the conditions under which the Rehm-Weller treatment applies, have been discussed (38) from a theoretical point of view. Models based on the Marcus treatment have been successfully applied to charge separation and transport in organic photoconductive layers used in electrophotography (qv), and have proven useful in the design of such systems (39).

An important consequence of the Marcus equation is the existence of the Marcus inverted region, which is not predicted by the Rehm-Weller equation. Thus, maximum electron-transfer rate occurs when $\Delta G^\circ = -\lambda$; if the reaction is more exoergic, the rate slows down in a mirror image of the relationship by which rate increases with increasing exoergicity up to the maximum, as shown in Figure 2. Experimental demonstration of the existence of the Marcus inverted regime (40) represented a validation of the Marcus model. Existence of the inverted region can be exploited in practical applications to control back-electron transfer (occurring with first-order rate constant, k_{et}) which competes with involvement of one or both of the charge separated products in the desired reaction (occurring with pseudo-first-order rate constant, k_p). Back-electron transfer is an energy wasting process (41).

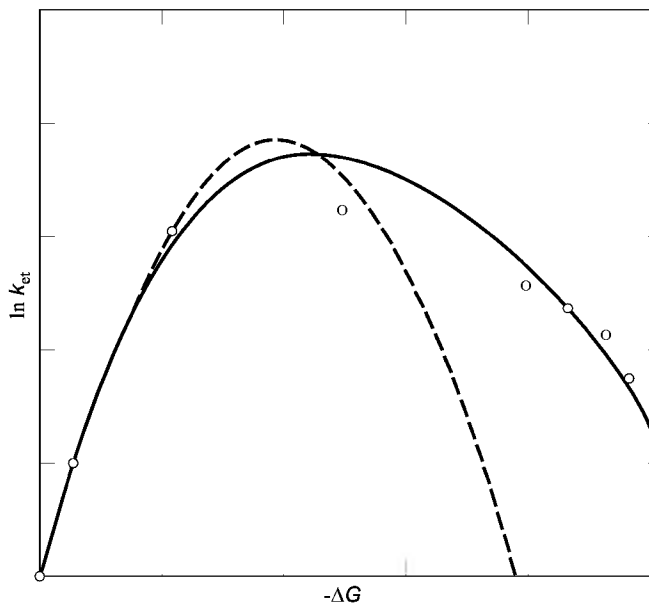


Fig. 2. Electron-transfer reaction rate, k_{et} , vs exoergicity of reaction: the dashed line is according to simple Marcus theory; the solid line and data points are experimentally determined (37).

Courtesy of The American Chemical Society.

Knowledge of photoinduced electron-transfer dynamics is important to technological applications. The quantum efficiency, ϕ , ie, the number of chemical events per number of photons absorbed of the desired electron-transfer photoreaction, reflects the competition between rate of the electron-transfer process, eg, from D^* , and the radiative and radiationless decay of the excited state, reflected in the lifetime, τ , of D^* in absence of A . Thus,

$$\phi = \frac{k[A]}{k[A] + \tau^{-1}} \left[\frac{1}{k_r + k_{et}} \right]$$

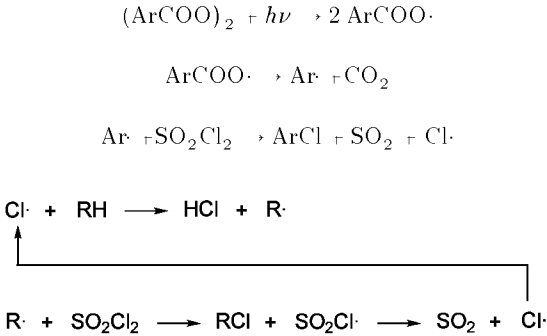
A strategy for increasing ϕ by inhibiting back-electron transfer, k_{et} , involves molecular design so as to maximize λ for the back reaction. For example, hydrazines undergo significant intramolecular reorganization associated with pyramidal-planar reconfiguration at nitrogen, on one-electron oxidation (42). *Ab initio* molecular orbital calculations have shown that this feature of the hydrazine functional group is preserved when it is incorporated into the pyrazoline ring structure (43). Thus triarylpyrazolines can be used as especially efficient electron-transfer sensitizers of decomposition of both arylidonium cations (44) for photoinitiation of free-radical and/or cationic polymerization processes. The theory of photoinduced electron-transfer dynamics is a tool to be used in the engineering of applied photochemical systems.

Applications

Photohalogenation. Photochemical chlorination of aliphatic hydrocarbons has been the textbook example of industrial photochemistry for decades (45). As of the mid-1990s it is still commercially important and industrial-scale halogenation has been reviewed in detail (1). In most examples of

historical importance, ultraviolet radiation was used to dissociate Cl₂ to yield atomic chlorine which, in turn, abstracts a hydrogen atom from the hydrocarbon substrate yielding an organic free radical and HCl. Recombination of this radical with atomic chlorine yields a monochlorocarbon which is more susceptible to the hydrogen abstraction reaction than the parent hydrocarbon. A mixture of chlorohydrocarbons usually results by this method (see CHLOROCARBONS AND CHLOROHYDROCARBONS). Production of trichloroethane occurs in this manner. Photohalogenation is also used for production of the insecticide lindane (γ-hexachlorocyclohexane) (see INSECT CONTROL TECHNOLOGY).

A more energy-efficient variation of photohalogenation, which has been used since the 1940s to produce chlorinated solvents, is the Kharasch process (45). Ultraviolet radiation is used to photocleave benzoyl peroxide (see PEROXIDES AND PEROXIDE COMPOUNDS). The radical products react with sulfuryl chloride (from SO₂ and Cl₂) to liberate atomic chlorine and initiate a radical chain process in which hydrocarbons become halogenated. Thus, for Ar = aryl,



The most innovative photohalogenation technology developed in the latter twentieth century is that for purposes of photochlorination of poly(vinyl chloride) (PVC). More highly chlorinated products of improved thermal stability, fire resistance, and rigidity are obtained. In production, the stepwise chlorination may be effected in liquid chlorine which serves both as solvent for the polymer and reagent (46). A solid-state process has also been devised in which a bed of microparticulate PVC is fluidized with Cl₂ gas and simultaneously irradiated (47). In both cases the reaction proceeds, counterintuitively, to introduce Cl exclusively at unchlorinated carbon atoms on the polymer backbone.

Laser Photochemical Vapor Deposition. Laser pyrolytic and laser photochemical vapor deposition (LPCVD) technologies based on pyrolysis and photolysis of organometallic precursors are important to the deposition of thin inorganic films and surface patterning for the microelectronics industry (2). Owing to use of organometallics this technology is sometimes called metal organic chemical vapor deposition (MOCVD). Thin films (qv) of metals, insulators, and semiconductors can be deposited on inert surfaces by decomposition of precursors from the gas phase. The chemistry can occur either in the gas phase adjacent to the substrate or in the adsorbed state on its surface (48,49). High resolution patterning of the surface, ie, imagewise deposition, requires that surface photochemistry via the gas-phase processes dominates; such imagewise deposition may be achieved either by writing the pattern using a scanned laser (50) or by imaging through a mask (51). A schematic diagram of a scanned laser imager for LPCVD application is shown in Figure 3 (48,50). Principal documented applications have involved repair of electronic devices, such as fabrication masks, interconnecting gate arrays and chips, and fabrication of customized circuits. Slow deposition rates appear to have been the principal limitation of LPCVD. The majority of applications have involved metal deposition. Metal alkyls are principal precursors of metal deposits, especially aluminum, in these processes. Only the best characterized materials deposited by LPCVD, ie, metals and insulating SiO_x (49), have been incorporated into commercial electronic products as of this writing.

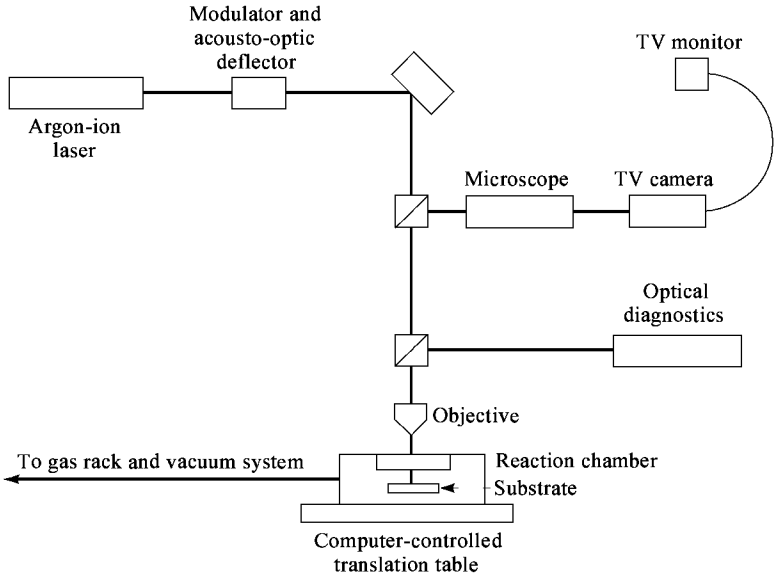


Fig. 3. Scanned laser imaging device for LPCVD in the microelectronics industry (48).

Courtesy of The American Chemical Society.

Distinction between pyrolytic and photochemical techniques is not clear cut. Deposition may, for example, involve thermolysis of a photochemically generated intermediate or vice versa. Photochemical techniques generally have an advantage over the pyrolytic approach in that it is not necessary to subject the substrate, even locally, to the extreme temperatures (750–1400°C) used in the purely pyrolytic processes (50). Frequency-doubled Ar ion lasers (244 nm) and excimer lasers (193 and 248 nm) are used to effect the photochemistry. Laser selection may be critical in a particular application. Where surface photochemistry is desired exclusively for high resolution patterning, decomposition of adsorbed Cd(CH₃)₂ or Zn(CH₃)₂ to yield the corresponding metal requires 193 nm photons from an ArF laser, even though the gas-phase decomposition of these precursors can be effected at 248 nm with a KrF laser. Surface energy relaxation pathways have been invoked to explain this difference. Metal carbonyls (qv) are also useful, especially for deposition of Ni, Fe, Cr, W, and Mo. These decompositions are often hybrid photochemical–pyrolytic (52). Noble or coinage metal films, Cu, Au, Ir, Pt, or Pd, are accessible by laser photolysis of volatile acetylacetonate (acac) or hexafluoroacetylacetonate (hfacac) complexes, again using excimer lasers. In these cases higher purity metal films are obtained by the pyrolytic, rather than the pure photochemical mechanism (48,53); surface temperatures of no more than 175–225°C are required to decompose hfacac complexes.

Excimer lamps have opened the possibility of cost-effective large-area direct photochemical vapor deposition (PCVD). PCVD of stoichiometric, insulating SiO₂ onto Si wafer has been reported using SiH₄ and N₂O as gas-phase precursors and the 172-nm radiation from a Xe*₂ lamp (54). Deposition

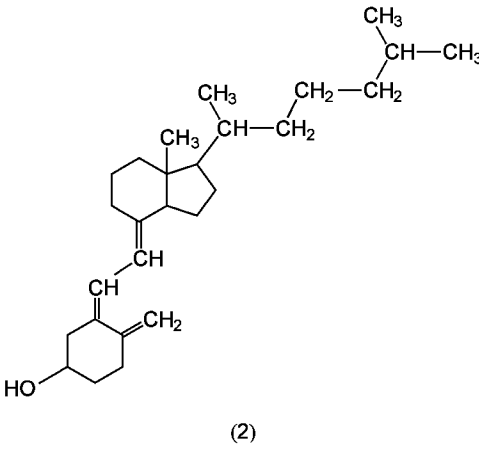
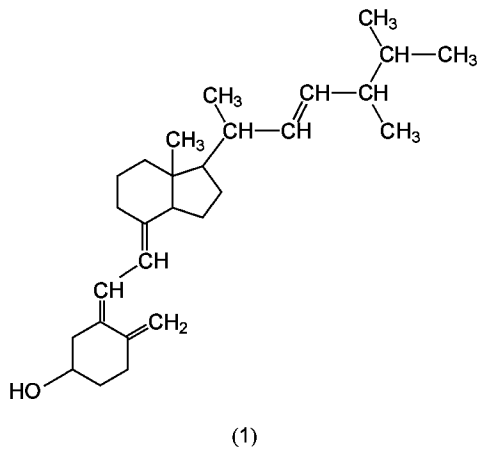
rate in this demonstration process is ca 3 nm min. Similarly α -Si_xC_{1-x} of useful electronic properties has been deposited on glass, also using the Xe*₂ lamp at a rate of 50 nm min, which is competitive with other CVD processes (55).

Epitaxial deposition of semiconductors (qv) by LPCVD is a promising technique for device fabrication, particularly radiation detectors, and is under intense investigation as of this writing. Studies on the deposition of crystalline, polycrystalline, or amorphous Si films have been motivated by thin-film solar cell applications. Decomposition of SiH₄ is usually pyrolytic because this common starting material does not absorb significantly beyond 160 nm; biphotonic excitation at high photon fluxes allows use of 193-nm radiation from an ArF laser to effect LPCVD of Si (56). This process is relatively inefficient. The preferred starting materials are Si₂H₆ or Si₃H₈, both of which undergo one-photon decomposition at 193 nm (57). GeH₄ absorbs at 248 nm (KrF laser) and is a useful precursor of Ge films by LPCVD; biphotonic excitation is thought to be involved, however (58). The photolytic technique is used for preparation of large-area Ge–Si alloy films from Si₂H₆–GeH₄ mixtures (59) and alternating amorphous Ge and Si structures can also be deposited photochemically for synthesis of multiquantum-well devices (60). By contrast, thermal chemical vapor depositions (CVD) of Ge yields primarily a crystalline Ge deposit (61).

Development of photochemical technology for epitaxial deposition of Group 12–16 (II–VI) semiconductors from organometallic precursors has focused on narrow band gap materials, eg, HgCdTe, CdTe, and HgTe, which are useful in near-infrared radiation detectors (62). The advantage of the photochemical technology over pyrolytic CVD is lower deposition temperature (63). MOCVD has also been used for deposition of Group 13–15 (III–V) semiconductors, eg, GaAs, InGaAs, and AlGaAs, but pyrolytic decomposition of the precursors is usually involved (49). LPCVD of the wider band gap III–V semiconductor GaN for application in blue light-emitting diodes (LEDs) has been investigated in the laboratory (64) and a hybrid photochemical–pyrolytic mechanism inferred. In addition to LPCVD of metallic layers, deposition of insulating SiO_x, primarily large-area deposition of SiO₂, has reached commercialization. As in the case of 172-nm excimer lamp photolysis, stoichiometric films result from 193-nm photooxidation of SiH₄ in the presence of N₂O (65). Photolysis of N₂O in the gas phase to form reactive oxygen atoms is most likely the primary process involved (66).

With the advent of organic optical polymers, an innovative photochemical approach to waveguide fabrication has been developed (67,68). Accordingly a chromophore, eg, a merocyanine dye, dissolved in or bound to a thin film of an optical polymer is imagewise photobleached to tailor the index of refraction (at nonactinic wavelengths) of the polymer in precise fashion. Channel waveguides (67) and active waveguide devices, eg, directional couplers and interferometric intensity modulators (68), can be fabricated by this technique. Lasers, eg, Ar ion, are usually employed to effect photobleaching having high spatial precision. Kinetic studies indicate that one-photon processes dominate (69).

Vitamin D. Vitamin D is synthesized by photochemical means (see VITAMINS). The commercial process was patented in 1985 (70). The term vitamin D is actually applied to several isomers wherein the side chain in the parent structure varies. Vitamin D₂ (1) and D₃ (2) are the most important, commercially, owing to use as animal feed additives. These compounds are synthesized from the steroids ergosterol and 7-dehydrocholesterol, respectively. The synthesis involves a photochemical ring opening (photo-Cope rearrangement) to yield the corresponding pre-vitamin D. The reaction is usually carried out in ether or alcohol using medium pressure Hg-vapor lamps which have higher output in the useful spectral regime (260–313 nm) than in the undesirable regime below 254 nm, as compared to low pressure lamps.



The *cis*-configuration of the 6,7-double bond in pre-vitamin D is critical to its subsequent thermal rearrangement to the active vitamin. A photochemical isomerization of pre-vitamin D to yield the inactive *trans*-isomer occurs under conditions of synthesis, and is especially detrimental if there is a significant short wavelength component, eg, 254 nm, to the radiation continuum used to effect the synthesis. This side reaction reduces overall yield of the process and limits conversion yields to ca 60% (71). Photochemical reconversion of the inactive side product, tachysterol, to pre-vitamin D allows recovery of the product which would otherwise be lost, and improves economics of the overall process (70).

Polymer-Based Technologies. Materials science aspects of polymer-based photochemical technologies have been reviewed (72). One example is the use of photoinitiated polymerization in three-dimensional engineering prototyping. In contemporary computer-aided design (CAD) a three-dimensional object is created on the computer. In this process the designer is guided by various digitally generated projections of the object on the monitor. In order to have a solid prototype for visual inspection of the final design, a photopolymerization technique, sometimes called stereolithography, is employed. Principles of this process were first described in the open literature in 1982 (73) (see COMPUTER AIDED DESIGN AND MANUFACTURING (CAD/CAM)).

As illustrated in Figure 4, the model is created on a platform submerged in a bath that is mounted on a turntable and contains a suitable monomer including a photoinitiator responsive to a laser line. The laser is modulated by the design computer and scanned radially while the turntable rotates to write

one laminar element of the design. The platform is then displaced downward by the thickness of one lamina, and the next layer created, and so forth. After the process is completed, the model is removed from the bath, washed free of unreacted monomer, and often subjected to a thermal or uv post-cure. Stereolithographic fabrication times of under one hour are commonplace. Annual sales of equipment and materials for this application now exceed \$30 million in the United States (74). Because polymerization involves formation of covalent bonds rather than intermolecular van der Waals interactions, photopolymerizable materials often shrink during curing. Shrinkage of the polymer is a primary limitation of the dimensional fidelity of stereolithography (75).

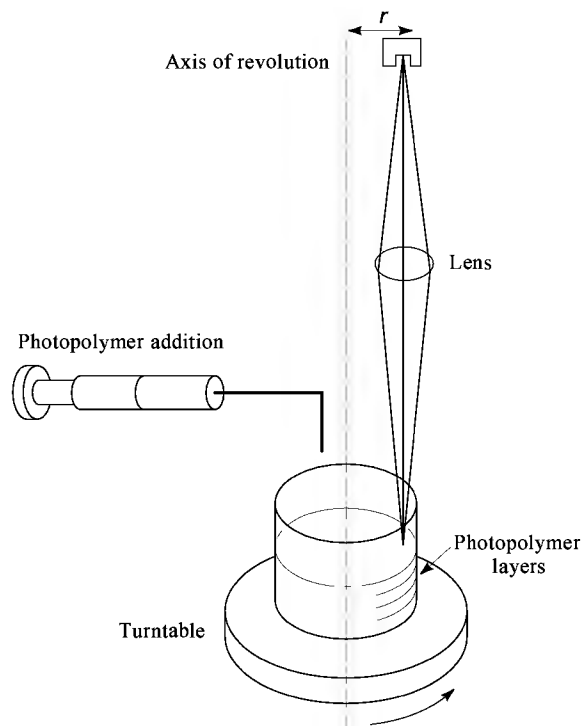
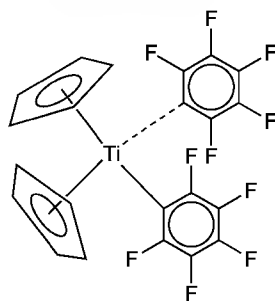


Fig. 4. Apparatus for stereolithographic generation of engineering prototypes in the CAD workflow (73).

Courtesy of The Society for Imaging Science and Technology (ISCT).

Thickness of the laminar layer is determined both by the need to reproduce fine detail in the object and by the penetration depth of the actinic laser light into the monomer bath (21,76). There is thus a trade-off between precision of detail in the model and time required for stereolithography, ie, the number of layers that have to be written, and an optimum light-absorbing initiator concentration in the monomer bath corresponding to the chosen layer thickness. Titanocene-based initiators, eg, bis-perfluorophenyltitanocene has been recommended for this application (77). Mechanistic aspects of the photochemistry of titanocenes and mechanisms of photoinitiation have been reviewed (76).



Photoinitiated polymerization can be used to replicate microstructured surfaces having a high degree of precision. An example is the Philips 2P Process (Fig. 5a) used in the fabrication of laser video disks (78). On a video disk, or compact audio disk (CD), the digitized information is recorded in the form of a spiral track of microscopic pits, typically ca 0.1- μm deep and 0.4- μm wide. Pit length and linear separations vary from 0.5–2.0 μm , and the pitch of the spiral is ca 1.0- μm between tracks (12,79). In the 2P process the information is initially formatted onto a Ni-plated master, called a stamper, which serves as a mold. Between the stamper and the disk substrate, made from optical quality poly(methylmethacrylate) (PMMA), is injected the curable monomer incorporating a photoinitiator. The monomer is cured in place to replicate the microstructure pattern of the stamper. The polymer replica is then metallized by vapor deposition, typically Al, to make it reflective. Video disks are double sided; that is, two metallized replicas are laminated together using an appropriate adhesive to make the finished product (see Fig. 5b). The stamper is made by conventional photolithographic technology using a novolak photoresist imaged using an Ar ion laser at 488 nm (see RESIST MATERIALS) (79).

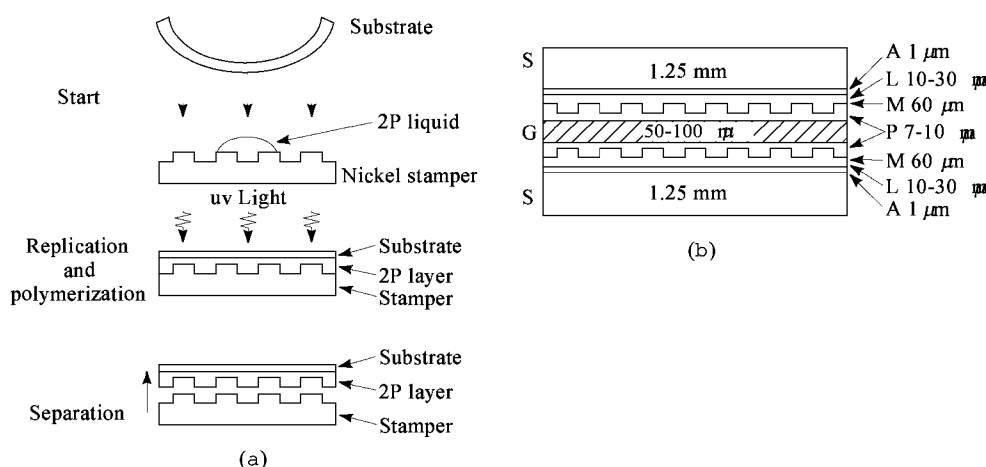
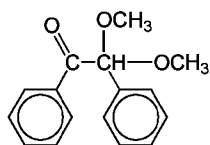


Fig. 5. (a) Schematic of the Philips 2P Process for video disk replication (79); (b) configuration of a double-sided LaserVision disk where the hole at the center is not shown. S represents the transparent substrate; A, the primer layer; L, lacquer with picture and sound information in the form of pits; M, mirror coating; and G, adhesive layer.

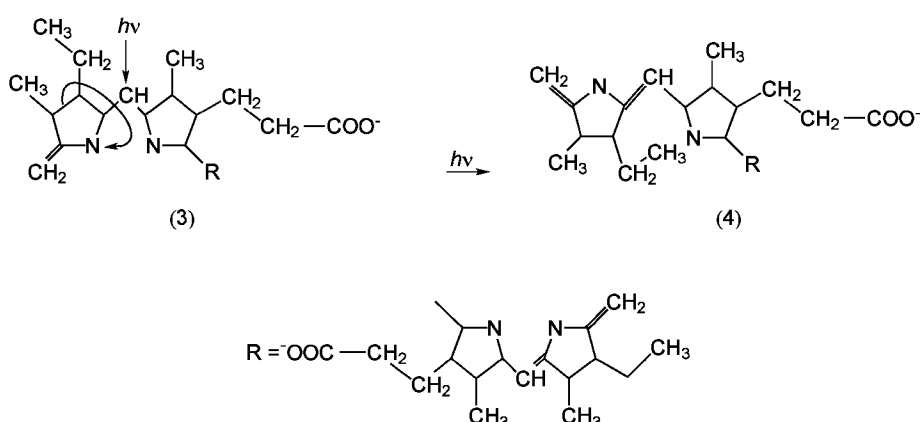
Courtesy of N.V. Philips' Gloeilampenfabrieken.

The photopolymer employed for the replication step must have extremely high dimensional stability, both during and after the curing process. In addition the light-cured polymeric replica must be hard, release easily from the stamper, and provide good adhesion to the metallic vapor coat (80). Free-radical polymerizable acrylic monomers, eg, a mixture of 2-ethylhexyl acrylate and 1,4-butanediol diacrylate, have been chosen by Philips for this application. The difunctional acrylate provides the cross-link density necessary to meet the hardness requirement for the cured polymer. The shrinkage problem is exacerbated if the cross-link density is too high, but a little shrinkage is tolerable insofar as it facilitates release of the replica from the stamper. Photoinitiation of polymerization of such mixtures is typically effected using a benzilketal-type initiator, eg, α,α -dimethoxy- α -phenylacetophenone (DMPA), responsive in the near-uv. Similar technology is employed to replicate the aspherical lenses used to focus the reading laser precisely on the pit tracks of the video disk or CD in their corresponding players (80).



Photochemical Therapies. The discipline of photomedicine has its origin (81) in observations, ca 1900, of the Danish Nobel laureate in Medicine, Niels Finsen, who recognized the photochemical basis of the disappearance during the summer of the facial lesions *lupus vulgaris*, Norwegian tuberculosis victims developed during the winter. Finsen eliminated *lupus vulgaris* by subjecting his Nordic patients to artificial light focused through a water-filled lens (to eliminate infrared) having quartz windows (to transmit uv).

A particularly successful therapeutic application of photochemistry has been the treatment of hyperbilirubinemia, otherwise known as newborn jaundice (82). Newborn jaundice, which occurs in nearly half of all newborn infants, results from high concentrations of the heme pigment, bilirubin, in the plasma, which can lead to irreversible brain damage if not controlled. The basis for photochemical treatment of newborn jaundice stems from the discovery in the 1930s (83) that bilirubin is efficiently decomposed when exposed to light. The therapeutic potential of this phenomenon resulted from the benchmark observation of an anonymous nurse that newborn jaundice disappeared in infants placed in a brightly sunlit portion of the hospital nursery (82,84). Devices for effecting whole-body irradiation of newborns are commercially available. These utilize cool white or blue fluorescent lamps, or appropriately filtered tungsten-halogen lamps (82). The key photochemical transformation of bilirubin is a geometric isomerization about a double bond converting a (Z,Z)-isomer (3) to an (E,Z)-isomer (4) of bilirubin. The latter isomer is nontoxic (85).



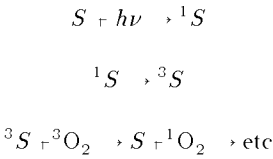
This is essentially the same process responsible for deactivation of photoexcited cyanine dyes.

New impetus was given to photomedicine by development of lasers that are compatible with the clinical environment. These include HeNe, Ar ion, ruby, and tunable dye lasers operating in the continuous wave (cw) mode. Prior to the advent of lasers in medicine, only the treatment of newborn jaundice, and the application of long wavelength uv irradiation in conjunction with administration (or topical application) of psoralen class sensitizers to treatment of skin diseases (86), principally psoriasis, were clinically important phototherapies.

Lasers can be coupled efficiently to fiber optic devices to deliver intense monochromatic light precisely to the desired region of the body, including internal organs (see FIBER OPTICS). As in other cases of laser-induced photochemistry, biphotonic effects may be important (87). Lasers also offer the advantage of being able to concentrate the incident energy in a spectral bandpass matched to the absorption band of the sensitizer.

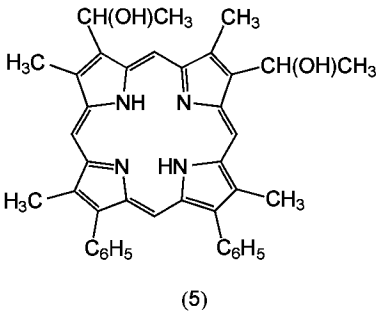
A principal application for photomedicine is the photodynamic treatment of cancer. Photochemical and clinical aspects of this topic have been reviewed (10,11). Direct irradiation of tumors coupled with administration of a sensitizer is used to effect necrosis of the malignancy. In this process, an excited state sensitizer interacts with dissolved O_2 *in vivo* to effect conversion of the oxygen from its triplet ground state to an excited singlet state, 1O_2 , which is highly cytotoxic. In principle, excited sensitizers in either the singlet or the triplet state can effect this conversion of molecular oxygen (8). In

practice interaction of excited sensitizer and O₂ is at best a diffusion controlled reaction and the solubility of O₂ in aqueous media is only ca 10⁻⁴ M. The Einstein-Smoluchowski relationship predicts that excited-state lifetimes of <1 μs are required for efficient sensitization, restricting useful sensitizers, S, to those giving high yields of longer-lived triplet excited states. This also implies that dyes characterized by high triplet yields are more apt to be skin sensitizers, as ¹O₂ is also thought to play an intermediate role in acute and chronic photodermatoses, and perhaps photocarcinogenesis (88).



As penetration of tissue by light increases more or less monotonically beyond 450 nm (89), it is desirable for sensitizer absorption bands and actinic laser lines to be located as far into the red as possible. Primarily sensitizers absorbing at ca 630 nm, matched to the 633-nm output of the HeNe laser, have been used. If the sensitizers are available, wavelengths out to 800 nm that are accessible using solid-state diode lasers might be preferable (11). Waste energy is converted to heat in tissue, and the power which can be delivered to the target area without causing thermal damage to surrounding, healthy tissue generally limits laser powers useful in photodynamic therapy. Clinical experience has shown steady-state irradiances of up to ca 200 J cm² at 633 nm to be tolerable (10). Higher powers would be tolerable at longer wavelengths, owing to the lower light absorption coefficient of the tissue. On the other hand, owing to photobleaching of typical sensitizers by the light actinic for tumor necrosis, the process is characterized by strong high intensity reciprocity failure, ie, an increase in light intensity is accompanied by a subproportional decrease, if any, in irradiation time required for the same biological effect (90).

A successful sensitizer for photodynamic therapy must therefore exhibit good optical absorption properties as far into the red as possible, matched to an available laser line, as well as efficient intersystem crossing to a long-lived triplet state. In addition, it must exhibit correct pharmacological properties, specifically persistence *in vivo* on the appropriate time scale, and selective absorption into the target tumor(s) (91). The latter minimizes sensitization of healthy tissue, as well as the amount of foreign substance which must be introduced into the body. Historically, a mixture of hematoporphyrin (5) derivatives was used, but higher molecular weight oligomers of hematoporphyrin which were by-products of the derivatization reaction were primarily responsible for tumor localization (92). Oligomeric hematoporphyrins are commercially available under the name Photofrin and comprise the most commonly used sensitizer for photodynamic cancer therapy. About 5000 patients had been treated successfully for various cancers using the photodynamic method and Photofrin as of 1992 (10). Selective absorption of oligomeric hematoporphyrins to malignant tissue combined with efficient uv-excited fluorescence of the adsorbed sensitizer has led to a proposal (11) that these same compounds could be used to enable early, noninvasive detection of malignancies, eg, lung tumors, by fluorescence.



Given stringent requirements for effective sensitizers and the desire to use wavelengths further to the red for therapeutic applications, definition of newer sensitizers has been a principal area of research since about 1987. Expanded theoretical and experimental understanding of photophysics has been a key element in identifying new classes of potential sensitizers (93–98). Research has focused on cationic derivatives of Nile Blue (93), metallo-phthalocyanines (94), naphthalocyanines (95), chlorin-type compounds (96), expanded ring porphyrinoids (97), as well as porphyrins other than hematoporphyrin and its derivatives (98). This work has also been reviewed (10,91). Instrumentation for photodynamic therapy has been reviewed (99).

BIBLIOGRAPHY

"Photochemical Technology" in *ECT* 2nd ed., Vol. 15, pp. 331–354, by C. R. Mullen, The Dow Chemical Co.; in *ECT* 3rd ed., Vol. 17, pp. 540–559, by J. J. Bloomfield and D. C. Owsley, Monsanto Co.

1. A. M. Braun, M.-T. Maurette, and E. Oliveros, in D. F. Ollis and N. Serpone, eds., *Photochemical Technology*, John Wiley and Sons, Ltd., Chichester, U.K., 1990.
2. V. Carassiti, *Coord. Chem. Revs.* **125**, 351 (1993).
3. N. Torimoto, *J. Chem. Educ.* **64**, 332 (1987); K. Honda, personal communication, 1992; D. Miller, *Indigo, from Seeds to Dye*, 3rd ed., Indigo Press, Aptos, Calif., 1984.
4. J.-L. Marignier, *Bull. Soc. Vaud. Sc. Nat.* **81**, 199 (1992); *La Recherche* **22**, 374 (1991); *Nature* **346**, 115 (1990).
5. J. P. Fouassier and J. F. Rabek, *Radiation Curing in Polymer Science and Technology*, Vols. I and II, Elsevier, London, 1993.
6. A. Reiser, *Photoreactive Polymers*, John Wiley and Sons, Inc., New York, 1989.
7. W. A. Hendrickson and M. C. Palazzotto, *Adv. Chem.* **238**, 411 (1993) and references cited therein. **Eur. Pat. 0,109,581** (1984), M. C. Palazzotto and W. A. Hendrickson; **U.S. Pat. 5,089,536** (1992), M. C. Palazzotto and W. A. Hendrickson (to 3M).
8. J. R. Norris, Jr., and D. Meisel, *Photochemical Energy Conversion*, Elsevier, New York, 1989, part I; M. R. V. Sahyun, *J. Photogr. Sci.* **31**, 243 (1983); F. Wilkinson and co-workers, *J. Am. Chem. Soc.* **115**, 12144 (1993).
9. J. D. Regan and J. A. Parrish, *The Science of Photomedicine*, Plenum Press, New York, 1982.
10. T. J. Dougherty, *Adv. Photochem.* **17**, 275 (1992).
11. S. L. Marcus, in C. J. Gomer, ed., *Future Directions and Applications in Photodynamic Therapy*, SPIE, Bellingham, Wash., 1990, pp. 5ff.
12. M. R. V. Sahyun, P. Vogelgesang, and J. M. Sturge, in V. Walworth, and A. Shepp, eds., *Imaging Materials and Processes*, 8th ed., Van Nostrand-Reinhold, New York, 1989, Chapt. 12.
13. R. Srinivasan, in J. M. Kelly, C. B. McArdle, and M. J. de F. Maunder, eds., *Photochemistry and Polymeric Systems*, Royal Society of Chemistry, Cambridge, U.K., 1992, pp. 47–53; G. A. Oldershaw, *Ibid.*, pp. 54–67.
14. Y. Tsuboi, K. Hatanaka, H. Fukumura, and H. Masuhara, *J. Phys. Chem.* **98**, 11237 (1994).
15. B. Eliasson and U. Kogelchatz, *Appl. Phys.* **B46**, 299 (1988).
16. H. Kumagai and M. Obara, *Appl. Phys. Lett.* **54**, 2619 (1989).
17. *Ibid.* **55**, 1583 (1984).
18. T. Hatakeyama, F. Kannari, and M. Obara, *Ibid.* **59**, 357 (1991).
19. **U.S. Pat. 5,233,762** (1993), R. Möller and J. Reiter (to Heidelberger Druckmaschinen AG).
20. **Eur. Pat. 0,604,738-A1** (1993), R. S. Nohr and J. G. MacDonald (to Kimberly-Clark Corp.).

21. G. A. Thommes and V. Weber, *J. Imaging Sci.* **29**, 112 (1985); A. R. Gutierrez and R. J. Cox, *Polym. Photochem.* **7**, 517 (1986).
22. R. J. DeVoe, P. M. Olofson, and M. R. V. Sahyun, *Adv. Photochem.* **17**, 313 (1992) and references cited therein.
23. A. W. Adamson, *J. Phys. Chem.* **71**, 798 (1967).
24. M. R. V. Sahyun and D. K. Sharma, *Chem. Phys. Lett.* **189**, 571 (1992).
25. J. C. Scaiano and L. J. Johnston, *Pure Appl. Chem.* **58**, 1273 (1986); J. C. Scaiano, L. J. Johnston, W. G. McGimpsey, and D. Weir, *Acc. Chem. Res.* **21**, 22 (1988); T. P. Davis, *J. Photochem. Photobiol., A: Chem.* **77**, 1 (1994) and references cited therein; R. J. Blankenau, G. L. Powell, W. P. Kelsey, and W. W. Barkmeier, *Lasers Surg. Med.* **11**, 471 (1991); C. Decker, *ACS Symp. Ser.* **266**, 207 (1984).
26. S. Speiser and N. Shakkour, *Appl. Phys. B* **38**, 191 (1985); S. Speiser, R. van der Werf, and J. Kommandeur, *Chem. Phys.* **1**, 297 (1973).
27. J. R. Lakowicz, I. Gryczynski, V. Bogdanov, and J. Kusba, *J. Phys. Chem.* **98**, 334 (1994).
28. P. D. Fleischauer, A. W. Adamson, and G. Sartori, *Inorganic Reaction Mechanisms*, Part II, John Wiley and Sons, Inc., New York, 1972.
29. J. E. Jones and P. B. Gilman, Jr., *Photogr. Sci. Eng.* **17**, 367 (1973); G. R. Bird, *ACS Symp. Ser.* **200**, 3 (1982) and references cited therein; T. Tani, *J. Imaging Sci.* **34**, 143 (1990); T. Tani, T. Suzumoto, and K. Ohzeki, *J. Phys. Chem.* **94**, 1298 (1990).
30. R. Humphry-Baker, M. Gratzel, and R. Steiger, *J. Am. Chem. Soc.* **102**, 847 (1980); T. Kunitake, *Ibid.* **104**, 4261 (1982).
31. **U.S. Pats. 2,756,227** and **2,734,900** (1953), D. W. Heseltine (to Eastman Kodak Co.); L. G. S. Brooker, in C. E. K. Mees and T. H. James, eds., *Theory of the Photographic Process*, 3rd ed., Macmillan Publishing, New York, 1966, Chapt. 11.
32. J. H. Pedstein and P. M. Borsenberger, *Extended Linear Chain Compounds*, Vol. 2, Plenum Press, New York, 1982, Chapt. 8.
33. D. F. Eaton, *Adv. Photochem.* **13**, 427 (1986); H.-J. Timpe, *Pure Appl. Chem.* **60**, 1033 (1988); I. C. Popovic, L. Katsikas, U. Möller, J. S. Velickovic, and H. Weller, *Macromol. Chem. Phys.* **195**, 889 (1994).
34. N. Serpone, R. Terzian, D. Lawless, and J.-M. Herrmann, *Adv. Electron Transfer Chem.* **3**, 33 (1994).
35. A. Weller, *Z. Phys. Chem., NF* **133**, 93 (1982); J. W. Verhoeven, *Pure Appl. Chem.* **58**, 1285 (1986).
36. D. Rehm and A. Weller, *Isr. J. Chem.* **8**, 259 (1970).
37. R. A. Marcus, *J. Phys. Chem.* **24**, 966 (1956); *Ann. Rev. Phys. Chem.* **15**, 155 (1964); *Electrochim. Acta* **13**, 995 (1968); V. G. Levich, *Physical Chemistry, an Advanced Treatise*, Vol. IX, Academic Press, Inc., New York, 1970, Chapt. 12.
38. N. Agmon and R. D. Levine, *Isr. J. Chem.* **19**, 230 (1980).
39. M. R. V. Sahyun, *Photogr. Sci. Eng.* **28**, 185 (1984); R. A. Albert, *J. Imaging Technol.* **10**, 64 (1983); H. Baumann, W. Helmstreit, and R. Ackermann, *J. Inf. Rec. Mater.* **19**, 29 (1991).
40. G. L. Closs and J. R. Miller, *Science* **240**, 440 (1988); J. R. Miller, L. T. Calcaterra, and G. L. Closs, *J. Am. Chem. Soc.* **106**, 3047 (1984).
41. I. R. Gould, D. Ege, S. L. Mattes, and S. Farid, *J. Am. Chem. Soc.* **109**, 3794 (1987); I. R. Gould, J. E. Moser, D. Ege, and S. Farid, *Ibid.* **110**, 1991 (1988).
42. S. F. Nelsen, *J. Am. Chem. Soc.* **110**, 6149 (1988).
43. J. T. Blair, M. R. V. Sahyun, and D. K. Sharma, *J. Photochem. Photobiol., A: Chem.* **77**, 133 (1994).
44. **Eur. Pat. 0,530,095-A1** (1992), M. R. V. Sahyun and G. P. Crooks (to 3M).
45. L. F. Fieser and M. Fieser, *Textbook of Organic Chemistry*, D. C. Heath, Boston, Mass., 1950.
46. **Belg. Pat. 889,968** (1981), and U.S. Pat. 4,377,459 (1983), R. G. Parker (to B. F. Goodrich, Inc.).
47. **Brit. Pat. 2,085,897** (1982), M. Balandier, J. Faurü, and C. Decker (to Rhône-Poulenc).
48. I. P. Herman, *Chem. Revs.* **89**, 1323 (1989); K. B. Ibbs and R. M. Osgood, Jr., *Laser Chemical Processing for Microelectronics*, Cambridge University Press, Cambridge, U.K., 1989.
49. J. G. Eden, *Photochemical Vapor Deposition*, John Wiley and Sons, Inc., New York, 1992.
50. G. D. Pazonis, H. Tang, and I. P. Herman, *IEEE J. Quantum Electron.* **25**, 976 (1989).
51. G. S. Higashi and C. G. Fleming, *Appl. Phys. Lett.* **48**, 1051 (1986).
52. H. H. Gilgen, T. Cacouris, P. S. Shaw, R. R. Krchnavek, and R. M. Osgood, Jr., *Appl. Phys. B* **42**, 55 (1987).
53. T. H. Baum, *J. Electrochem. Soc.* **134**, 2616 (1987).
54. P. Bergonzo, U. Kogelschatz, and I. W. Boyd, *Appl. Surf. Sci.* **69**, 393 (1993).
55. C. Manfredotti, F. Fizzotti, M. Boero, and G. Piatti, *Appl. Surf. Sci.* **69**, 127 (1993).
56. C. Fuchs, E. Boch, E. Fogarassy, B. Aka, and P. Siffert, *Mater. Res. Soc. Symp. Proc.* **101**, 361 (1988).
57. Y. Toyoshima, K. Kumata, U. Itoh, and A. Matsuda, *Appl. Phys. Lett.* **51**, 1925 (1987).
58. J. F. Osmundsen, C. C. Abele, and J. G. Eden, *J. Appl. Phys.* **57**, 2921 (1985).
59. H. H. Burke, I. P. Herman, V. Tavitian, and J. G. Eden, *Appl. Phys. Lett.* **55**, 253 (1989).
60. D. H. Lowndes and co-workers, *Appl. Phys. Lett.* **52**, 1868 (1988).
61. D. Eres and co-workers, *Mater. Res. Soc. Symp. Proc.* **131**, 517 (1989).
62. S. J. C. Irvine and co-workers, *J. Cryst. Growth* **93**, 732 (1988).
63. J. E. Jensen, P. D. Brewer, G. L. Olson, L. W. Tutt, and J. J. Zinck, *J. Vac. Sci. Technol. A* **6**, 2808 (1988).
64. P. C. John, J. J. Alwan, and J. G. Eden, *Thin Solid Films* **218**, 75 (1992).
65. P. K. Boyer, K. A. Emery, H. Zarnani, and G. J. Collins, *Appl. Phys. Lett.* **45**, 979 (1984).
66. K. F. Jensen, *Chemical Vapor Deposition*, Academic Press, Inc., New York, 1990.
67. S. Aramake, G. Assanto, and G. I. Stegemann, *Electron. Lett.* **26**, 1300 (1990); K. B. Rochford, R. Zanon, Q. Gong, and G. I. Stegemann, *Appl. Phys. Lett.* **55**, 1161 (1989); K. W. Beeson, K. A. Horn, M. J. McFarland, and J. T. Yardley, *Ibid.* **58**, 1955 (1991); A. E. Willner, M. N. Ruberto, D. J. Blumenthal, D. W. Podsebnik, and R. M. Osgood, Jr., *Ibid.* **54**, 1839 (1989).
68. E. Van Tomme, P. Van Daele, R. G. Baets, G. R. Mohlmann, and M. B. J. Diemeer, *J. Appl. Phys.* **69**, 6273 (1991); M. B. J. Diemeer and co-workers, *Electron. Lett.* **26**, 379 (1990).
69. R. S. Moshrefzadeh, D. K. Miesmer, M. D. Radcliffe, C. V. Francis, and S. Mohapatra, *Appl. Phys. Lett.* **62**, 16 (1993).
70. **U.S. Pat. 4,551,214** (1985), K.-H. Pfoertner and H.-J. Hansen (to Hoffmann-LaRoche).
71. H. J. C. Jacobs and E. Havinga, *Adv. Photochem.* **11**, 305 (1979).
72. A. Ledwith, in Ref. 13, p. 1ff.
73. A. J. Herbert, *J. Appl. Photogr. Eng.* **8**, 185 (1982); H. Kadama, *Rev. Sci. Inst.* **52**, 1770 (1981); **U.S. Pat. 4,575,330** (1985), C. W. Hull (to 3D Systems, Inc.).
74. *3rd Quarterly Report to Shareholders*, 3D Systems, Inc., Valencia, Calif., 1993.
75. J. Arline, *Rapid Prototyping Rep.* **1**, 7 (1991); C. Decker, in P. K. T. Oldring, ed., *Chemistry and Technology of UV and EB Formulation for Coatings, Inks and Paints*, Vol. 5, SITA Technologies, London, 1994, pp. 221–225 and references cited therein; J. S. Arney, L. Sanders, and R. Hardesty, *J. Imaging Sci. Technol.* **38**, 262 (1994).
76. A. Roloff, *Adv. Chem.* **238**, 399 (1993).
77. H. Angerer, V. Desobry, M. Riediker, H. Spahni, and M. Rembold, *Proceedings of the Conference on Radiation Curing ASLA*, Society of Manufacturing Engineers, Dearborn, Mich., 1988; **Eur. Pat. 0,122,223** (1984) and **U.S. Pats. 4,590,287** (1986), **4,910,121** (1990), and **4,962,012** (1990), M. Riediker, M. Roth, N. Bühler, and J. Berger (to Ciba-Geigy, Ltd.).
78. H. C. H. van Rijsewijk, P. E. J. Legierse, and G. E. Thomas, *Philips Tech. Rev.* **40**, 287 (1982).
79. G. Bouwhuis, *Principles of Optical Disc Systems*, Adam Hilger, Bristol, U.K., 1985.
80. J. G. Kloosterboer, *Adv. Polym. Sci.* **84**, 1 (1988); J. G. Kloosterboer, G. J. M. Lippits, and H. C. Meinders, *Philips Tech. Rev.* **40**, 298 (1982).
81. A. Hollaender, in Ref. 9, p. xiii.

82. T. R. C. Sisson and T. P. Vogl, in Ref. 9, Chapt. 16.
83. H. Fischer and K. Herle, *Hoppe-Zeylers Z. Physiol. Chem.* **251**, 85 (1938).
84. R. J. Cremer, P. W. Perryman, and D. H. Richards, *Biochem. J.* **66**, 600 (1957).
85. A. F. McDonagh, D. A. Lightner, and A. Wooldridge, *J. Chem. Soc., Chem. Comm.* **3**, 110 (1978).
86. J. A. Parrish, R. S. Stern, M. A. Pathak, and T. B. Fitzpatrick, in Ref. 9, Chapt. 18 and references cited therein.
87. D. E. Rounds, in Ref. 9, Chapt. 18.
88. H. H. Seliger and W. D. McElroy, *Light: Physical and Biological Action*, Academic Press, Inc., New York, 1965.
89. B. C. Wilson, in D. R. Doiron and C. J. Gomer, eds., *Porphyrin Localization and Treatment of Tumors*, Alan R. Liss, New York, 1984.
90. W. R. Potter, *Proc. SPIE* **1065**, 88 (1989); L. O. Svaasand, C. J. Gomer, and E. Morinelli, in Ref. 10, p. 233ff.
91. A. R. Morgan and D. Skalkos, in Ref. 11, p. 87ff.
92. S. Schwartz and T. J. Dougherty, *Photochem. Photobiol.* **38**, 377 (1983).
93. L. Cincotta, J. W. Foley, and A. H. Cincotta, *Photochem. Photobiol.* **46**, 751 (1987).
94. H. Brasseur, H. Ali, R. Langlois, and J. E. van Lier, *Photochem. Photobiol.* **46**, 739 (1987).
95. P. A. Firey and M. A. J. Rogers, *Photochem. Photobiol.* **45**, 535 (1987).
96. A. R. Morgan and N. C. Tertel, *J. Org. Chem.* **51**, 1347 (1986); A. M. Richter and co-workers, *J. Natl. Cancer Inst.* **79**, 1327 (1987); R. K. Pandey, D. A. Bellnier, K. M. Smith, and T. J. Dougherty, *Photochem. Photobiol.* **53**, 65 (1991).
97. G. Scherzman and co-workers, *Photochem. Photobiol.* **52**, 741 (1990).
98. M. C. Berenbaum and co-workers, *Br. J. Cancer* **54**, 717 (1986); K. Morris and A. D. Ward, *Tetrahedron Lett.* **29**, 2501 (1988).
99. G. Wagnieres and co-workers, in Ref. 11, p. 249ff.

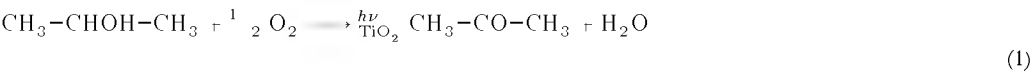
General References

A. Gilbert and J. E. Baggott, *Essentials of Molecular Photochemistry*, CRC Press, Boca Raton, Fla., 1991.
J. Michl, *Electronic Aspects of Organic Photochemistry*, John Wiley and Sons, Inc., New York, 1990.
J. D. Scaiano, *Handbook of Organic Photochemistry*, CRC Press, Boca Raton, Fla., 1989.
N. J. Turro, *Modern Molecular Photochemistry*, University Science Books, 1991.

M. R. V. Sahyun
3M Center

PHOTOCATALYSIS

Catalysis (qv) refers to a process by which a substance (the catalyst) accelerates an otherwise thermodynamically favored but kinetically slow reaction and the catalyst is fully regenerated at the end of each catalytic cycle (1). When photons are also implicated in the process, photocatalysis is defined without the implication of some special or specific mechanism as the acceleration of the prate of a photoreaction by the presence of a catalyst. The catalyst may accelerate the photoreaction by interaction with a substrate either in its ground state or in its excited state and or with the primary photoproduct, depending on the mechanism of the photoreaction (2). Therefore, the nondescriptive term photocatalysis is a general label to indicate that light and some substance, the catalyst or the initiator, are necessary entities to influence a reaction (3,4). The process must be shown to be truly catalytic by some acceptable and attainable parameter. Reaction 1, in which the titanium dioxide serves as a catalyst, may be taken as both a photocatalytic oxidation and a photocatalytic dehydrogenation (5).



Photochemical processes provide an alternative to the traditional U.S. EPA recognized routes to water (qv) purification: air stripping (removal of volatile components) and carbon adsorption (removal of both volatile and nonvolatile contaminants) (6). A weakness of these traditional processes is that pollutants are not destroyed, rather the pollutants are moved from one phase to another. Ultraviolet (uv) light in combination with oxidation processes can remove bacterial substances and dissolved organics from solution. Advanced oxidation processes (AOPs), such as uv O₃, uv H₂O₂, and heterogeneous photocatalytic methods, result in relatively rapid and complete destruction of numerous organics, including halogenated hydrocarbons (see CHLOROCARBONS AND CHLOROHYDROCARBONS). Heterogeneous photocatalysis uses air or oxygen, rather than O₃ or H₂O₂, and a semiconductor photocatalyst at low (ambient) temperature (see SEMICONDUCTORS) (7). This process leads to total mineralization of organic pollutants to CO₂ without significant formation of photocyclized intermediate products. Also, the photocatalyst employed, eg, TiO₂ or ZnO, is inexpensive and can be supported on suitable materials (6,8) (see CATALYSTS, SUPPORTED). This technology has industrial potential (9–11). Research in photocatalysis has been driven by legislation in industrialized countries encouraging water decontamination and simultaneous contaminant destruction (12).

Advanced Oxidation Processes

Homogeneous Photocatalysis. Remediation of wastewaters from organic contaminants presupposes some oxidative process to burn off the dissolved carbonaceous species utilizing such oxidizers as ozone (qv) or hydrogen peroxide (qv). It is important, however, that the oxidative process lead to total destruction of the contaminants. Used alone, neither oxidizing agent is efficient. The simultaneous coupling of light and an oxidant, such as O₃, H₂O₂, TiO₂, or others, however, has led to mineralization of organic carbon to CO₂ (9–12). Figure 1 depicts the absorption spectra of H₂O₂ and O₃ along with the terrestrial air mass one (AM1) solar emission spectrum (13–15). These AOPs unmistakably implicate some organic radical species generated by photolysis or by assisted photolysis (reaction with ·OH radicals) of the organic substrates. Subsequent trapping by molecular oxygen (air) yields peroxy radicals and peroxides along their way to complete mineralization.

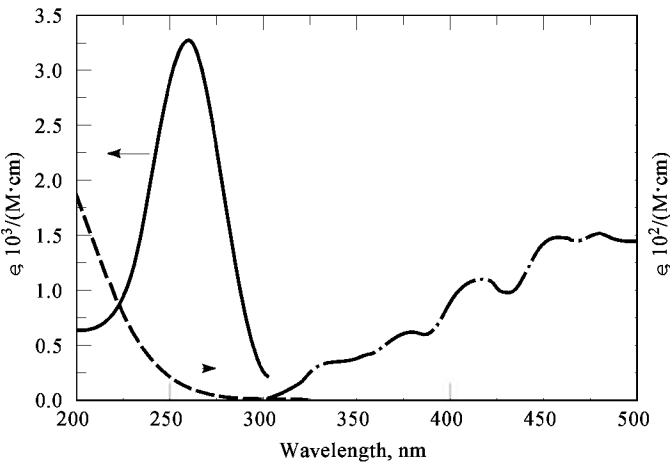


Fig. 1. Absorption spectra of (—) H₂O₂, and (---) O₃ in aqueous media together with the (— · —) terrestrial air mass one (AM1) solar emission spectrum in arbitrary units. The extinction coefficient, ε, of ozone is over 10 times greater than that of H₂O₂ (15).

Growth in the development of AOPs (9–11) devoted to water and air purification has led to various review articles germane to photocatalysis (16,17). The common thread is the presence and action of the hydroxyl radical, ·OH. The steps implicated in two AOPs, O₃ uv and H₂O₂ uv, are summarized in Figure 2. Two other systems, O₃ H₂O₂ dark and O₃ H₂O₂ uv light are also shown (13,14). The H₂O₂ uv system is considerably less effective in uv light utilization than is O₃ uv (Fig. 1) as evidenced by the large variation in the molar extinction coefficients at 254 nm; ε_{H₂O₂} = 19 (M⁻¹cm), ε_{O₃} = 2850–3000 (M⁻¹cm) (14). The former uses little if any of the uv portion of AM1 sunlight. However, although O₃ uv is more efficient in using light energy and quantitatively yields H₂O₂ (Fig. 2) (18,19), only ca 5% of the photolyzed ozone yields ·OH radicals (14). In solutions which contain organic substrates likely to absorb uv light, H₂O₂ becomes rather inefficient unless large concentrations are used (13). Formation of ·OH is more efficient in alkaline (pH > 7) than in acidic media for O₃ HO₂·. This finds equivalence with O₃ H₂O₂ dark which necessitates pHs > 7 for production of ·OH radicals and oxidation of organics. At pHs > 7, there are no advantages in using O₃ uv over O₃ H₂O₂ dark for generating ·OH radicals (14). A frequent disadvantage in alkaline aqueous solutions is the ubiquitous presence of CO₃²⁻, and or HCO₃⁻, depending on pH. The CO₃²⁻ is a better scavenger of ·OH radicals by two to three orders of magnitude than is HCO₃⁻ to produce CO₃^{·-}. Thus solutions having pH below ca 6 are preferred. Where CO₃^{·-} and ·OH radicals coexist, the latter radical is the dominant oxidizing species.

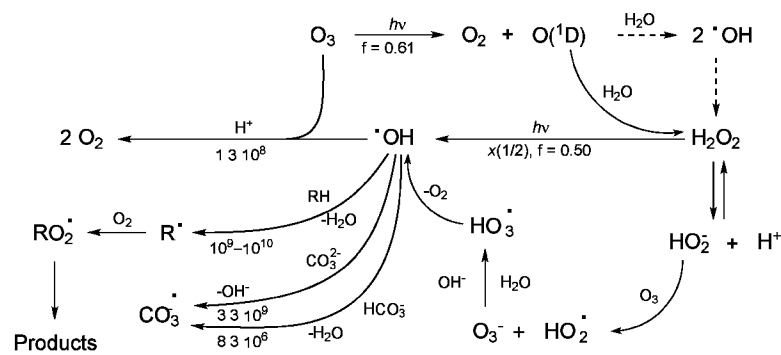


Fig. 2. Steps in advanced oxidation process (AOPs) involving ozone, hydrogen peroxide, and uv light of 254 nm. (¹D) represents the doublet state; ϕ represents quantum yield, and the other numbers associated with the reaction arrows are rate constants in units of (M s)⁻¹. Dashed arrows indicate processes less likely to occur.

The more efficient system of generating $\cdot\text{OH}$ radicals in the homogeneous phase is H_2O_2 uv, where the quantum yield, $\phi_{254\text{nm}}$ is 0.50 (20,21). There are some disadvantages in high flow rate reactors, however, where short residence times of the organic substrate(s) to be oxidized in contact with the oxidizing species require high rates of $\cdot\text{OH}$ radical production (14). Formation of $\cdot\text{OH}$ radicals from H_2O_2 uv is independent of pH and temperature (13), but its practical efficiency does depend on pH and on the quantities of carbonates present in wastewaters. Figure 3 shows this dependence for the total organic carbon (TOC) decay in the photooxidation of a landfill leachate (22). It is evident that the process is more efficient in acidic media.

The AOP processes, O_3 uv, H_2O_2 uv, and TiO_2 uv are compared in Figure 4 (22). Figure 4a shows the percent TOC attenuation versus irradiation time for three AOP combinations in the photooxidative mineralization of a nitroxyene test molecule (23). The H_2O_2 uv process has little effect, whereas the TiO_2 air uv leads to more efficient decay. Addition of H_2O_2 to give the TiO_2 H_2O_2 air uv system makes the process even more efficient. However, after one hour of irradiation, both the TiO_2 air uv and the TiO_2 H_2O_2 air uv achieve the same degree of mineralization. Figure 4b compares the H_2O_2 , O_3 , and O_3 H_2O_2 dark systems to uv radiation and combinations of oxidant and uv in the degradation of a landfill leachate. Both O_3 H_2O_2 uv and H_2O_2 uv show good efficiency. Under otherwise identical conditions, the TiO_2 uv system, not shown in Figure 4b, is somewhat less efficient in detoxifying leachate than these other AOPs unless H_2O_2 is added.

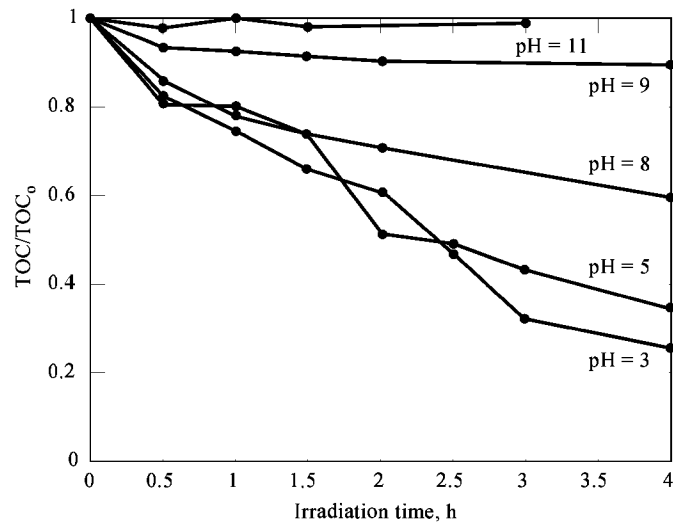


Fig. 3. pH Dependence of total organic carbon (TOC) decay during the photooxidation of a landfill leachate using H_2O_2 uv. Initial TOC concentration, TOC_0 , is 1260 mg L (22).

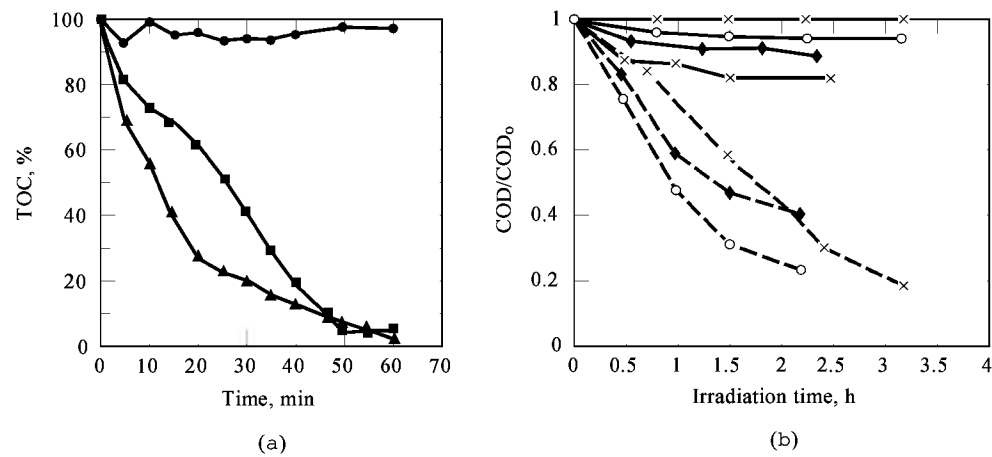
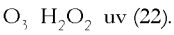


Fig. 4. (a) TOC decrease as a function of illumination time for the photodegradation of 1,2-dimethyl-3-nitrobenzene [83-41-0] in aqueous media in an annular reactor where the initial TOC = 45 mg L; temperature, 45°C; and wavelength of irradiation >320 nm. The (●) represents H_2O_2 uv; (■), TiO_2 uv in air; and (▲), TiO_2 H_2O_2 uv. (b) Decay of chemical oxygen demand (COD) for a landfill leachate at pH 3 as a function of irradiation time where (x) represents the presence of H_2O_2 ; (○), the use of uv radiation; and (◆), O_3 . For the dashed lines, (x) represents H_2O_2 uv; (◆), O_3 uv; and (○), O_3 H_2O_2 uv.



Heterogeneous Photocatalysis. Heterogeneous photocatalysis is a technology based on the irradiation of a semiconductor (SC) photocatalyst, for example, titanium dioxide [13463-67-7], TiO_2 , zinc oxide [1314-13-2], ZnO , or cadmium sulfide [1306-23-6], CdS . Semiconductor materials have electrical conductivity properties between those of metals and insulators, and have narrow energy gaps (band gap) between the filled valence band and the conduction band (see ELECTRONIC MATERIALS; SEMICONDUCTORS).

A number of electronic and photochemical processes occur following band gap excitation of a semiconductor. Figure 5 illustrates a sequence of photochemical and photophysical events and the possible redox reactions which might occur at the surface of the SC particle in contact with a solution. Absorption of light energy greater than or equal to the band gap (E_{bg}) of the semiconductor results in a shift of electrons from the valence band (VB) to the conduction band (CB) and the creation of holes (h^+) in the valence band (eq. 2). These charge carriers



recombine, radiatively and/or nonradiatively, in competition with rapid diffusion to the surface where the resulting nonequilibrium distribution of electrons and holes gives rise to reduction or oxidation processes of adsorbed species, surface groups, and the SC components. When illuminated with light of sufficient energy, the SC particle becomes part of a particulate system capable of mimicking a microphotoelectrochemical cell at which efficient reductive and oxidative processes may take place (24). The efficiency of these processes is largely determined by five factors (25): (1) efficient absorption of solar light with minimal entropy production; (2) fast charge separation after light absorption; (3) separation of products in order to prevent reverse reactions; (4) adjustment of the redox potentials of the excited states to the redox reactions which store the energy; and (5) long-term stability or continuous reproduction. The dominant competitive process to charge separation is $e^- - h^+$ recombination, which for semiconductor clusters primarily occur via nonradiative processes (26).

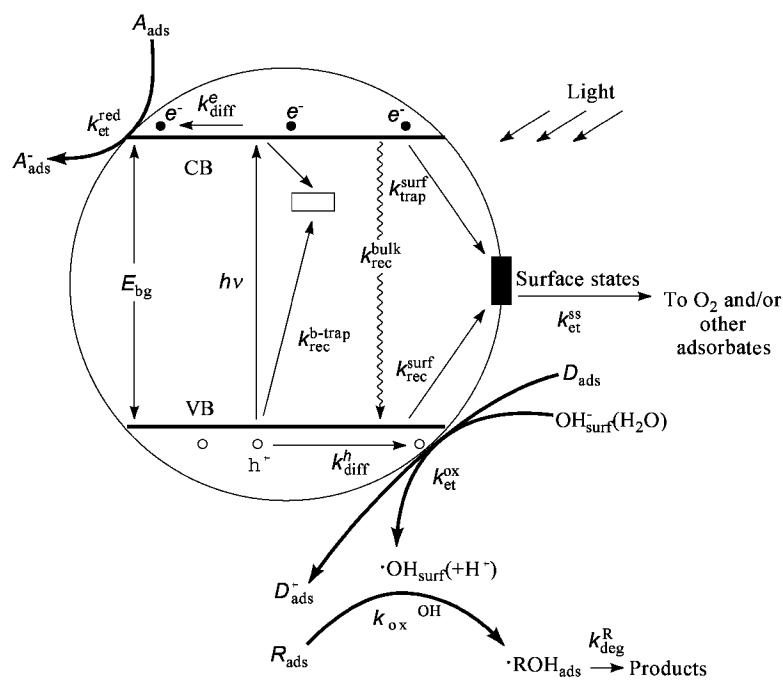


Fig. 5. Photophysical and photochemical processes in a semiconductor cluster where A_{ads} and D_{ads} represent chemical species, adsorbed on the surface of the semiconductor particle, which are capable of undergoing reduction (A_{ads}) and oxidation (D_{ads}) at rates $k_{\text{et}}^{\text{red}}$ and $k_{\text{et}}^{\text{ox}}$, respectively. The subscript et = electron transfer. Other reaction rates are designated according to the species undergoing the reaction, eg, $k_{\text{et}}^{\text{OH}}$. The subscripts diff, deg, ox, rec, surf, and trap refer to the processes of diffusion, degradation, oxidation, recombination, surface, and trapping, respectively. E_{bg} = band gap energy; CB and VB refer to the semiconductor conduction band and valence band, respectively. See text.

Figure 6 illustrates band gaps and band edge positions in aqueous media (26,27) for a number of SC materials which have been explored in heterogeneous photocatalysis. Also shown are the regions of redox potential for the oxidation of organic groups to illustrate the thermodynamic limitations of the type of photoreactions that can be carried out with the photogenerated electrons and holes. Thus, reduction of a given species (A) (Fig. 5) occurs if the redox level of A lies below the CB level of the semiconductor. Similarly, if oxidation of D is to be carried out, its redox level must be positioned above the valence band of the semiconductor. When the redox level of A (or D) is located between the valence and conduction bands of the semiconductor, both reduction and oxidation processes can occur.

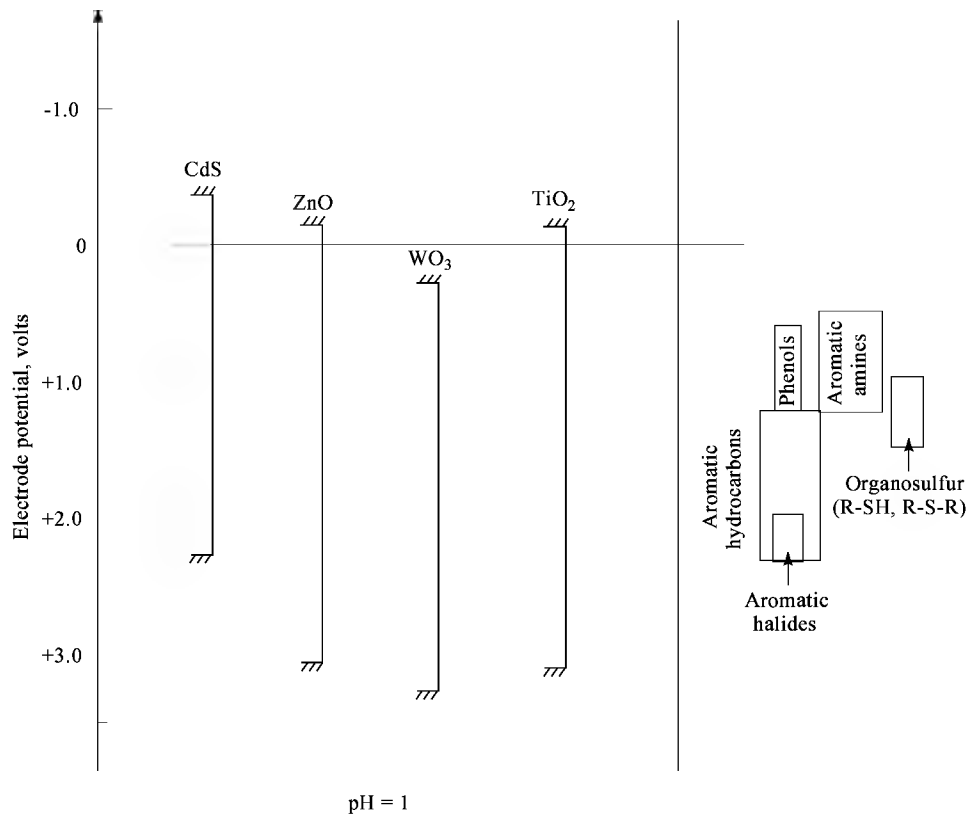


Fig. 6. Band edge positions of several semiconductors in contact with an aqueous electrolyte at pH 1 in relation to the redox (electrode) potential regions (vs the standard hydrogen electrode) for the oxidation of organic functional groups (26,27).

Titanium dioxide has been used extensively to photocatalyze the mineralization of a large number of organic pollutants in aqueous dispersions (16) and is the photocatalyst discussed herein. This photocatalyzed mineralization typically proceeds via formation of a series of intermediates of progressively higher oxygen-to-carbon ratios which eventually are oxidized quantitatively to CO₂ and H₂O. In the case of phenol [108-95-2], a model compound often used in heterogeneous photocatalysis studies, mineralization proceeds via formation of several hydroxylated intermediates that include predominantly catechol and hydroquinone (Fig. 7) (28). Organic compounds containing phosphorus, sulfur, and nitrogen atoms are oxidized quantitatively to PO₄³⁻, SO₄²⁻, and NO₃⁻, respectively, in addition to CO₂ (29); halocarbons yield X⁻ ions. In some cases, N atoms are converted reductively to NH₃ (30).

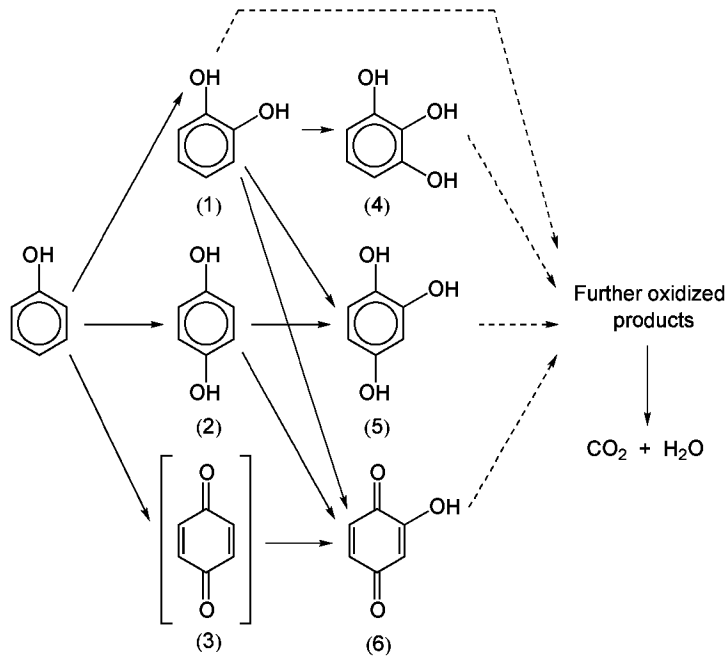


Fig. 7. Photocatalyzed degradation of phenol showing principal intermediates including catechol [120-80-9] (1), hydroquinone [123-31-9] (2), *p*-benzoquinone [106-51-4] (3), pyrogallol [87-66-1] (4), hydroxyhydroquinone [533-73-3] (5), and hydroxybenzoquinone [2474-72-8] (6). The dashed lines indicate processes less likely to occur.

Photooxidations. Some of the compounds which have been successfully photocatalytically degraded are listed in Table 1 (16). These compounds include chlorinated aromatics, surfactants (qv), pesticides (qv), and herbicides (qv). The total mineralization of long-chain alkanes and such alkyl derivatives as dodecane [112-40-3], dodecyl sulfate, 1-bromododecane, and decanoic acid [143-07-7] has been demonstrated (31). The photocatalyzed destruction of simple and complex chlorinated derivatives of alkanes, alkenes, carboxylic acids, and aromatics has also been achieved. In addition to those compounds given in Table 1, examples include 4-chlorophenol, 2,4,5-trichlorophenol [39399-44-5], and chlorobenzene [108-90-7]. Although chlorinated compounds have been more extensively studied, there are examples involving brominated, eg, bromoform [75-25-2], and fluorinated compounds as well as compounds containing phosphorus, sulfur, and nitrogen (16).

Table 1. Photocatalytically Degraded Organic Pollutants

Compound	CAS Registry Number	Structure
pentachlorophenol	[87-86-5]	
<i>m</i> -fluorophenol	[372-20-3]	
2,4,5-trichloro-phenoxyacetic acid	[93-76-5]	
4,4'-dichlorodiphenyl-trichloroethane (4,4'-DDT)	[50-29-3]	
3,3'-dichlorobiphenyl (3,3'-DCB)	[2050-67-1]	
2,7-dichlorodibenzo- <i>p</i> -dioxin	[33857-26-0]	
6-chloro- <i>N</i> -ethyl- <i>N'</i> -(1-methylethyl)-1,3,5-triazin e-2,4-diamine (atrazine)	[1912-24-9]	

Studies carried out on anionic, cationic, and nonionic surfactants have shown that the aromatic and hydrophilic portions of molecules are easily oxidized, whereas the long hydrocarbon chains are converted at slower rates. Surfactant activity does, however, disappear upon loss of the aromatic portion, thereby reducing the nuisance of the reactants (32). Total mineralization to CO₂ has been demonstrated for nonionic polyethoxylated 4-nonylphenols having average numbers of 2,5, and 12 ethoxy units (33).

The herbicide bentazon [25057-89-0] (3-(isopropyl-2,1,3-benzothiadiazin-4-one-2,2-dioxide) is nearly quantitatively mineralized to yield SO₄²⁻ and CO₂ (34). The *s*-triazine herbicides, eg, atrazine, simazine [122-34-9], trietazine [1912-26-1], prometon [1610-18-0], and prometryn [7287-19-6], are rapidly degraded but are resistant to total mineralization. The final product in these cases, in addition to NO₃⁻, SO₄²⁻, and Cl⁻ ions, is cyanuric acid [108-80-5], a very stable six-membered ring -(N=C(OH))_3 (35) of low toxicity (36). The only compound that is resistant to initial photocatalyzed oxidative attack is CCl₄ (37).

Water Treatment. Several components must be treated simultaneously in a multicomponent mixture as available in wastewaters to prove the technology of heterogeneous photocatalysis. The formation and subsequent elimination of intermediates in the photooxidative process must be monitored, identifying all intermediates and final products.

Actual water treatment challenges are multicomponent. For example, contamination of groundwater by creosote [8021-39-4], a wood (qv) preservative, is a recurring problem in the vicinity of wood-preserving facilities. Creosote is a complex mixture of 85 wt % polycyclic aromatic hydrocarbons (PAHs); 10 wt % phenolic compounds, including methylated phenols; and the remaining 5 wt % N-, S-, and O- heterocyclics (38). Aqueous solutions of creosote are therefore, in many ways, typical of the multicomponent samples found in polluted aquifers.

In the case of the mineralization of creosote (39), light alone, or the semiconductor catalyst TiO₂ alone, have little if any effect on transformation. However, light plus titania yield a relatively rapid transformation of creosote into intermediate products, which ultimately are decomposed to carbon dioxide (Fig. 8). The formation of stoichiometric quantities of CO₂ takes longer than the disappearance of the individual components of creosote because intermediates form that also must be degraded. Hence, mineralization invariably involves multicomponent systems. Total mineralization necessitates demonstration of stoichiometric evolution of the ultimate oxidation product, CO₂. The experiment was carried out in a saturated oxygen atmosphere in a septum-sealed Pyrex glass reactor. Total mineralization took about six to seven hours.

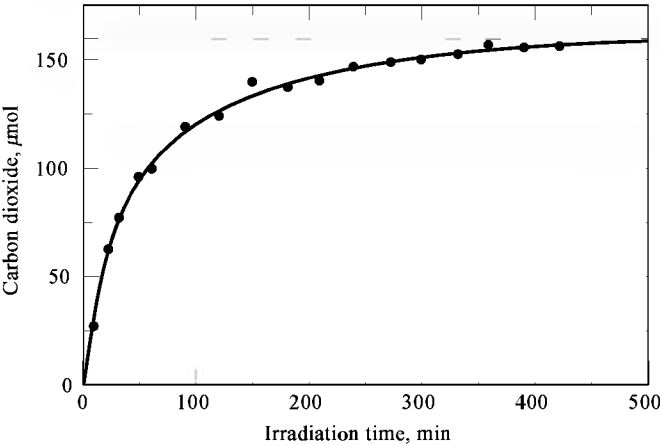


Fig. 8. The evolution of stoichiometric quantities of carbon dioxide formed from the total photomineralization of 100 ppm of creosote in water under a saturated oxygen atmosphere. A 25-mL sample having 75.5 wt % C was used. The dashed line corresponds to the expected stoichiometric quantity of CO₂.

Mechanism of Heterogeneous Photocatalysis

A debate centers on the mechanistic details of heterogeneous photocatalysis. The goal is to improve the photocatalytic activity of TiO₂, and understand the role and importance of mineralization by (1) free versus surface bound oxidizing radicals, ·OH, and (2) by surface ·OH radicals versus direct hole oxidation.

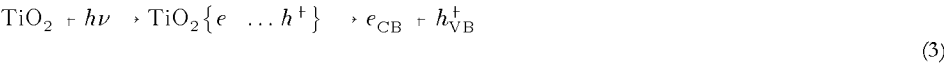
Two principal pathways have been established in mineralization of organic substrates and oxidation of inorganic materials, eg, CN⁻. One considers surface OH groups or H₂O on TiO₂ as the primary target(s) for the reaction of photogenerated holes, a reaction which yields ·OH radicals. The prevailing view favors these radicals as primary oxidizing species. The alternative route implicates direct hole oxidation of the organic substrate (40). Another problem is whether the primary oxidation event of ·OH radicals occurs on the surface of the photocatalyst or whether ·OH desorb and react with oxidizable substrates in solution.

Titanium Dioxide Particle Surface. In the dark and in a given aqueous electrolyte medium, the titania surface has certain electronic characteristics and possesses a distinct number of adsorption sites onto which anions, cations, organics, and other species present can chemisorb or physisorb, reversibly or irreversibly. There exist two types of surface OH groups: an OH group that bridges two surface vicinal Ti⁴⁺ ions (a Brnsted acid center), and a terminal Ti⁴⁺–OH group, which possesses basic character. The surface also contains chemisorbed and physisorbed water of hydration. The number of surface OH groups is ca 5 nm², representing less than 50% of surface coverage; theoretically the number is ca 5–10 nm². The exact number is dependent on the type of crystal plane (41). The consensus seems to be 7–10 OH nm² (or 7–10 × 10¹⁴ OH cm²) for TiO₂ at ambient temperature. Chemisorbed water molecules bound directly to surface Ti⁴⁺ ions amounts to about 2–3 nm² for rutile TiO₂. Most if not all the Ti⁴⁺ sites are occupied.

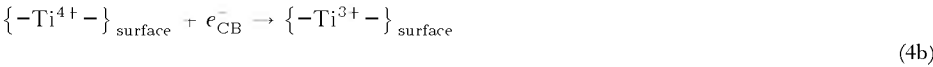
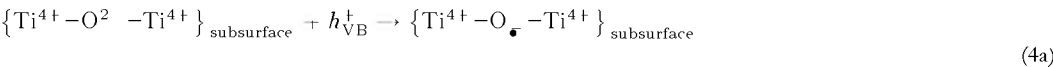
Other species present in the medium can chemisorb strongly (and irreversibly) on surface sites, eg, H₂PO₄⁻, HPO₄²⁻, F⁻, and NO₃⁻, displacing some of the terminal OH groups to give a surface coverage approaching half a monolayer (41). Other anions, eg, Cl⁻, I⁻, and SO₄²⁻, are reversibly adsorbed. This extrinsic adsorption has important consequences in photocatalyzed oxidations, because anions can potentially block catalytic sites and scavenge redox equivalents. The nonstoichiometry of TiO₂ provides surface Ti³⁺ sites on which electron acceptors such as molecular oxygen can adsorb. High (400–600°C) temperature pretreatment of TiO₂ in an oxidizing (O₂) atmosphere reduces the number of these reductive sites; pretreatment in a reducing (H₂) atmosphere increases their number.

Under Irradiation. Under illumination, the surface characteristics can undergo dramatic changes, altering the nature of the adsorption sites. Thus, dark adsorption–desorption events are altered, and additional events arising from photoadsorption–photodesorption equilibria take place.

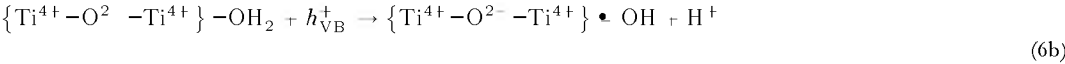
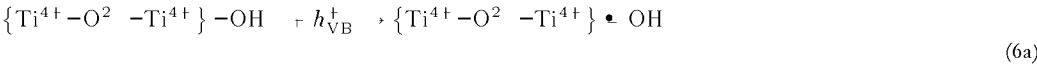
The primary photochemical act, subsequent to near-uv light (wavelengths <400 nm) absorption by TiO₂ particles, is generation of electron–hole pairs where the separation (eq. 3) into conduction band electrons (e_{CB}⁻) and valence band holes (h_{VB}⁺) is facilitated by the electric field gradient in the space charge region. Chemically, the hole associated with valence band levels is constrained at



the surface or subsurface sites in the region where light is absorbed. The great mobility of the electron (42), poised at the conduction band potential, facilitates its migration across the particle. Both charge carriers migrate rapidly to the surface in a few picoseconds, where they are ultimately trapped by intrinsic subsurface energy traps {Ti⁴⁺–O²⁻–Ti⁴⁺} for the hole and surface traps {–Ti⁴⁺–} for the electrons (eq. 4) (43), and by extrinsic traps via interfacial electron transfer with surface adsorbed electron donors (D_{ads}) and acceptors (A_{ads}), respectively (eq. 5) (see Fig. 5).

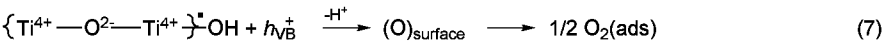


Rapid e⁻–h⁺ recombination, the reverse of equation 3, necessitates that D and A be pre-adsorbed prior to light excitation of the TiO₂ photocatalyst. In the case of a hydrated and hydroxylated TiO₂ anatase surface, hole trapping by interfacial electron transfer occurs via equation 6 to give surface-bound ·OH radicals (43,44). The necessity for pre-adsorbed D and A for efficient charge carrier trapping calls attention to the importance of adsorption–desorption equilibria in

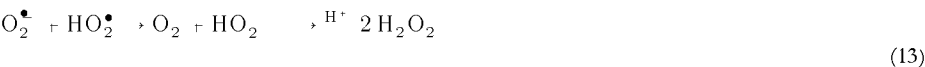


photocatalysis. These equilibria and the extent of adsorption depend on the pH of the medium and the point of zero charge (pzc) for the semiconductor used. For anatase TiO₂, pzc is ca 6.0–6.4 (45). In turn this is highly affected by the particle environment, ie, nature of ions and ionic strength, among others. In acid media, the particle surface is positively charged and thus should enhance adsorption (qv) of anionic and polar substrates; in alkaline media the surface charge is negative and should favor adsorption of cationic species.

Trapped e⁻ and h⁺ rapidly recombine on the particle surface. The lifetime of h⁺ is <30 ns (46). To obviate recombination, e⁻s are scavenged by pre- and photoadsorbed molecular oxygen to give superoxide radical anions, O₂^{•-} (ads), which can be reduced further to the peroxide dianion, O₂²⁻ (ads). Surface peroxo species can form (47) either by hydroxyl radical (hole) pairing or by sequential two-hole capture by the same OH group (eqs. 6a and 7) or by dismutation of O₂^{•-} (eq. 8).



In acidic media, ie, pH < 3, $\text{O}_2^{\bullet -}$ protonates to give $\text{HO}_2^{\bullet}$ ($\text{p}K_a$ 4.88 (48)). Other reactions that occur on the TiO_2 particle surface and which are solvent (water) related are summarized in equations 9 to 13.



The extent of the changes in the electronic characteristics and in the very nature of the TiO_2 particle surface dictates the events that take place along the photo oxidative path to mineralization of organic substrates.

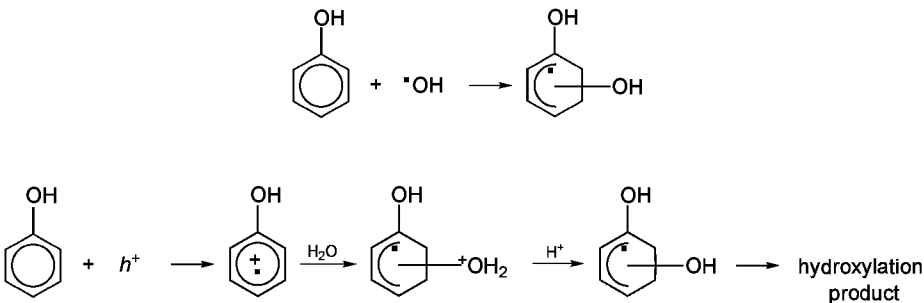
The Oxidizing Species. Chemical evidence favors $\bullet\text{OH}$ radicals as the principal oxidizing species in the photomineralization of most organics examined (49). This evidence originates from observations that hydroxylated intermediate products, formed along the course of the photooxidation process, bear close resemblance to products obtained by oxidation using Fenton's reagent (50). In the oxidation of phenol by light-activated TiO_2 in aqueous dispersions, the products shown in Figure 7 were identified (28). The primary components were hydroquinone (HQ) and catechol (CC) for 16% conversion. Similar hydroxylated species, eg, 3-fluorocatechol, fluorohydroquinone, 4-fluorocatechol, and 1,2,4-trihydroxybenzene, were identified in the heterogeneous photocatalyzed oxidation of 3-fluorophenol (51). Added support for $\bullet\text{OH}$ as the primary oxidant in aqueous media comes from kinetic deuterium isotope experiments (52).

The nature of the intermediates implicated in the photooxidation of water with TiO_2 has been identified in several reports using spin traps by the electron spin resonance (esr) technique under ambient conditions (53). No evidence for $\bullet\text{OH}$ species, even at 4.2 K, was found (43), but the esr signal associated with the $\text{O}_2^{\bullet -}$ radical anion resulting from trapping positive holes at lattice oxide ions was inferred (eq. 4a). It appears that the $\bullet\text{OH}$ radical identified by spin trapping methods is not the primary product of hole trapping, but rather would seem to originate as a transient intermediate of photooxidation.

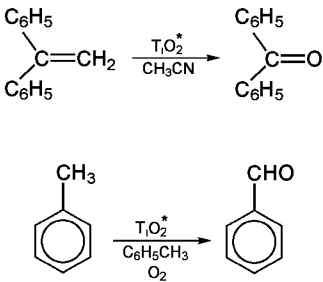
No clear picture of the primary radical intermediate(s) in the TiO_2 photooxidation of water has appeared. The nature of the observed radical species depends on the origin and pretreatment of the TiO_2 sample, on the conditions and extent of its reduction, on the extent of surface hydroxylation, and on the presence of adventitious electron acceptors such as molecular oxygen (41). The hole is trapped on the terminal OH group (54).

In competition experiments the presence of inhibitors, eg, $\bullet\text{OH}$ radical scavengers, and identification of intermediate products and their ratios have led to inferences that photooxidation of organics over light-activated TiO_2 (and ZnO) involves both $\text{OH}^{\bullet}$ radicals and photoholes, h^+ (55–59). Oxidation of acetate at illuminated TiO_2 –water interfaces produced the expected photo-Kolbe products CO_2 and methyl radicals by photoholes, and glycolate and glyoxylate by oxidation of acetate with $\bullet\text{OH}$ radicals via H-atom abstraction at the methyl group (57). The quantitative inhibition by ethanol, a $\bullet\text{OH}$ scavenger, of the photooxidation of dichlorobenzene over aqueous ZnO dispersions generated various hydroxylated intermediates (59). This is as evidence that the $\text{OH}^{\bullet}$ species is the sole oxidant. By contrast, the photooxidation of furfuryl alcohol (58) and monochlorophenols (56) has been suggested to proceed by both pathways: for the latter, about 65% via oxidation with $\text{OH}^{\bullet}$ radicals and the remainder ca 35% by direct hole oxidation.

Product identification does not distinguish $\bullet\text{OH}$ versus hole oxidation, because the products are identical. For example, the products identified in the photo oxidation of phenol (qv) (Fig. 7) may originate either by $\bullet\text{OH}$ radical attack of the phenol ring, or by direct hole oxidation to give the cation radical which subsequently undergoes hydration in solvent water.



Complete destruction of organics by an oxidative path over light-activated aqueous TiO_2 suspensions (TiO_2^*) does not occur if either H_2O and/or molecular O_2 are absent (60). Partial oxidations, and not mineralization, are the rule when photooxidations are carried out in redox-inert solvents such as acetonitrile and dichloromethane (61,62). In the absence of water, mineralization of the organic substrate does not occur, only partially oxidized products, often involving photooxygenation, have been isolated (62). In these cases the primary oxidizing species may be the photohole, h^+ , but the intervention of $\bullet\text{OH}$ species is by no means precluded. Reactions 14 and 15 illustrate some relevant examples (61,62).



Surface vs Solution Reactions. Another issue of debate in photocatalyzed mineralization of organic substrates is whether the initial oxidation occurs on the photocatalyst's surface or in solution. Kinetic data of photooxidations and photoreductions have often been fitted to the simple

rate, r_{initial} , expression shown in equation 16 (63–65):

$$r_{\text{initial}} = \frac{kKC}{1 + KC}$$

(16)

where k is the observed rate constant, K is the adsorption coefficient, and C is the initial concentration of substrate; or its more complex form, for a multicomponent system (eq. 17) (63) fit to a Langmuir adsorption isotherm (66),

$$r_{\text{initial}} = \frac{dC}{dt} = \frac{kKC}{\left\{ 1 + \sum_{\text{(all reactants)}} K_i C_i(t) + \sum_{\text{(all intermediates)}} K_j C_j(t) + K_S C_S \right\}}$$

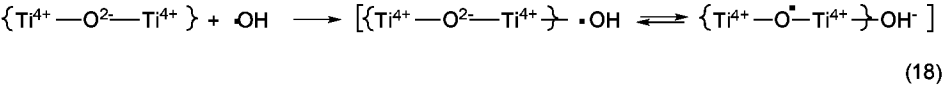
(17)

Inferences that oxidation takes place on the photocatalyst's surface have been made (67). No such conclusions can be drawn. Similar observations have been made in homogeneous media if a bimolecular reaction between two reactants is assumed. A Langmuir-type behavior is no guarantee of a surface occurring process. A rigorous treatment (68) of the kinetics involved in the photocatalyzed oxidations of organic substrates on an irradiated semiconductor has confirmed this.

Other studies have sought chemical evidence to ascertain whether or not the oxidation is a surface process. The selective inhibiting influence of isopropanol in the oxidation of furfuryl alcohol by ·OH radicals over ZnO dispersions suggests a homogeneous phase process (58). From the relative importance of the formation of glycolate and glyoxylate via OH· oxidation of acetate, which increases with increasing pH (57), it was inferred that little adsorption of acetate takes place on the negatively charged TiO₂, surface in alkaline pH, and that the hydroxyl radicals had to diffuse away from the surface of the photocatalyst to oxidize acetate in solution. Tenuous kinetic arguments based on apparent adsorption coefficients, K , have also been presented (69) to infer that, at high phenol surface coverage of the TiO₂ particles, the ·OH radical reacts at the surface, whereas at low coverage the ·OH freely diffuses into the solution where it contributes significantly to the overall photooxidation. A photoelectrochemical study (70) was strongly supportive of a solution ·OH reactive species in photocatalyzed oxidations. However, an esr study (71) concluded that the ·OH radical does all its work on the catalyst's surface, and photooxidation is a surface process.

Further indications that the ·OH radical is surface bound, and is unlikely to desorb into the solution, emanates from a study (72) which noted that decafluorobiphenyl [434-90-2] (DFBP) is tenaciously adsorbed (>99%) on metal oxide (Al₂O₃ and TiO₂) particle surfaces and does not undergo facile (<5%) exchange between the two oxide materials. When adsorbed on the alumina surface in dispersions into which H₂O₂ or a TiO₂ colloidal sol (titania particle size ca 0.05 μm) was added, followed by uv irradiation, the DFBP photodegraded. This indicated that the ·OH radical from H₂O₂ and TiO₂ sols (titania particles adsorbed on alumina) migrate to the reaction site on the DFBP–Al₂O₃ system to initiate the photooxidative events. By contrast, if TiO₂ beads (size ca 1000 μm) were used in lieu of H₂O₂ or the TiO₂ sol to generate the oxidizing species, the photodegradation was nearly suppressed and was identical to the behavior of the DFBP–Al₂O₃ system alone. Pentafluorophenol, which readily exchanges between the two metal oxide surfaces, underwent facile photodegradation under otherwise similar conditions. Thus, the photogenerated oxidizing species, ie, the ·OH radical, does not migrate far from the photogenerated active sites on TiO₂, and the degradation process must occur at the photocatalyst surface or within a few atomic distances from the surface (73).

Pulse radiolysis results (74) have led other workers to conclude that adsorbed ·OH radicals (surface trapped holes) are the principal oxidants, whereas free hydroxyl radicals probably play a minor role, if any. Because the OH· radical reacts with TiO₂ at a diffusion controlled rate, the reverse reaction, that is desorption of ·OH to the solution, seems highly unlikely. The surface trapped hole, as defined by equation 18, accounts for most of the observations which had previously led to the suggestion of ·OH radical oxidation. The formation of H₂O₂ and the observations of hydroxylated intermediate products could all occur via



surface reaction of this species. The possibility that a small fraction of the ·OH radicals may leak out from the surface and mediate the photooxidation process in solution cannot be precluded entirely.

BIBLIOGRAPHY

"Photochemical Technology" in *ECT* 2nd ed., Vol. 15, pp. 331–354, by C. R. Mullen, The Dow Chemical Co.; in *ECT* 3rd ed., Vol. 17, pp. 540–559, by J. J. Bloomfield and D. C. Owsley, Monsanto Co.

1. H. Kisch, in N. Serpone and E. Pelizzetti, eds., *Photocatalysis—Fundamentals and Applications*, Wiley-Interscience, New York, 1989, Chapt. 1.
2. C. Kotal, *Adv. Chem. Ser.* **238**, 1 (1993).
3. F. Chanon and M. Chanon, in Ref. 1, Chapt. 15.
4. N. Serpone, E. Pelizzetti, and H. Hidaka, in Z. W. Tian and Yi Cao, eds., *Photochemical and Photoelectrochemical Conversion and Storage of Solar Energy*, International Academic Publishers, Beijing, China, 1993.
5. S. J. Teichner and M. Formenti, in M. Schiavello, ed., *Photoelectrochemistry, Photocatalysis and Photoreactors*, Reidel Publishing Co., Dordrecht, the Netherlands, 1985, pp. 457–489.
6. N. Serpone, E. Borgarello, R. Harris, P. Cahill, M. Borgarello, and E. Pelizzetti, *Sol. Energy Mater.* **14**, 121 (1986).
7. D. F. Ollis, in M. Schiavello, ed., *Photocatalysis and Environment: Trends and Applications*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1988, p. 663, and references therein.
8. H. Al-Ekabi and N. Serpone, *J. Phys. Chem.* **92**, 5726 (1988).
9. G. R. Helz, R. G. Zepp, and D. G. Crosby, eds., *Aquatic and Surface Photochemistry*, Lewis Publishers, Boca Raton, Fla., 1994.
10. D. F. Ollis and H. Al-Ekabi, eds., *Photocatalytic Purification and Treatment of Water and Air*, Elsevier, New York, 1993.
11. *Proceedings of the First International Conference on Advanced Oxidation Technologies*, London, Ontario, Canada, June 25–July 1, 1994.
12. D. F. Ollis, E. Pelizzetti, and N. Serpone, *Environ. Sci. Technol.* **25**, 1522 (1991), and references therein.
13. J. R. Bolton and S. R. Cater, in Ref. 9, Chapt. 33, pp. 467–490.
14. C. C. David Yao and T. Mill, in Ref. 9, Chapt. 35, pp. 499–515.
15. Oriol Corp. Catalog, Vol. II, Stratford, Conn., 1994, pp. 1–215.
16. Bahnmann, J. Cunningham, M. A. Fox, E. Pelizzetti, P. Pichat, and N. Serpone, in Ref. 5, pp. 261–316, and references therein.
17. O. Legrini, E. Oliveros, and A. M. Braun, *Chem. Rev.* **93**, 671 (1993).
18. G. R. Peyton and W. H. Glaze, *ACS Symp. Ser.* **327**, 6 (1987).
19. H. Taube, *Trans. Faraday Soc.* **53**, 656 (1956).
20. J. P. Hunt and H. Taube, *J. Am. Chem. Soc.* **74**, 5999 (1952).
21. J. L. Weeks and M. S. Matheson, *J. Am. Chem. Soc.* **78**, 1273 (1956).

22. D. Weichgrebe, A. Vogelpohl, D. Bockelmann, and D. Bahnemann, in Ref. 10, pp. 579–584.

23. L. Jakob, E. Oliveros, O. Legrini, and A. M. Braun, in Ref. 10, pp. 511–532.

24. N. Serpone, E. Borgarello, and E. Pelizzetti, *J. Electrochem. Soc.* **135**, 2760 (1988).

25. H. Gerischer, *Pure Appl. Chem.* **52**, 2649 (1980).

26. M. Gratzel, *Heterogenous Photochemical Electron Transfer*, CRC Press, Boca Raton, Fla., 1988.

27. M. R. Rifi and F. H. Covitz, *Introduction to Organic Electrochemistry*, Marcel Dekker, Inc., New York, 1974.

28. K. Okamoto, Y. Yamamoto, H. Tanaka, M. Tanaka, and A. Itaya, *Bull. Chem. Soc. Jpn.* **58**, 2015 (1985); K. Okamoto, Y. Yamamoto, H. Tanaka, and A. Itaya, *Bull. Chem. Soc. Jpn.* **58**, 2023 (1985).

29. G. K.-C. Low, S. R. McEvoy, and R. W. Matthews, *Environ. Sci. Technol.* **25**, 460 (1991); G. K.-C. Low, S. R. McEvoy, and R. W. Matthews, *Chemosphere* **19**, 611 (1989).

30. K. Nohara, H. Hidaka, E. Pelizzetti, and N. Serpone, in press.

31. E. Pelizzetti and co-workers, *Ann. Chim.* **80**, 81 (1990).

32. H. Hidaka, H. Kubota, M. Gratzel, N. Serpone, and E. Pelizzetti, *Nouv. J. Chim.* **9**, 67 (1985); H. Hidaka, H. Kubota, M. Gratzel, E. Pelizzetti, and N. Serpone, *J. Photochem.* **35**, 219 (1986); H. Hidaka, Y. Fujita, K. Ihara, S. Yamada, K. Suzuki, and N. Serpone, *J. Jpn. Oil Chem. Soc.* **36**, 386 (1987); H. Hidaka, E. Pelizzetti, and co-workers, *J. Photochem. Photobiol. A: Chem.* **42**, 375 (1988).

33. E. Pelizzetti, C. Minero, V. Maurino, A. Sclafani, H. Hidaka, and N. Serpone, *Environ. Sci. Technol.* **23**, 1385 (1989).

34. E. Pelizzetti, V. Maurino, C. Minero, O. Zerbini, and E. Borgarello, *Chemosphere* **18**, 1437 (1989).

35. E. Pelizzetti, V. Maurino, C. Minero, V. Carlin, E. Pramauro, O. Zerbini, and M. L. Tosato, *Environ. Sci. Technol.* **24**, 1559 (1990).

36. E. Canelli, *Am. J. Public Health* **64**, 155 (1974).

37. D. F. Ollis, *Environ. Sci. Technol.* **19**, 480 (1985), and references therein.

38. J. G. Mueller, P. J. Chapman, and P. H. Pritchard, *Environ. Sci. Technol.* **23**, 1197 (1989).

39. N. Serpone, R. Terzian, D. Lawless, A.-M. Pelletier, C. Minero, and E. Pelizzetti, in Ref. 9, pp. 387–398.

40. R. B. Draper and M. A. Fox, *J. Phys. Chem.* **94**, 4628 (1990).

41. J. Augustynski, *Struct. Bonding* **69**, 1 (1988).

42. J. M. Warman, M. P. de Haas, N. Serpone, and P. Pichat, *J. Phys. Chem.* **95**, 8858 (1991).

43. R. F. Howe and M. Gratzel, *J. Phys. Chem.* **91**, 3906 (1987).

44. D. W. Bahnemann, A. Henglein, and L. Spanhel, *Faraday Discuss. Chem. Soc.* **78**, 151 (1984).

45. Y. G. Berube and P. L. de Bruyn, *J. Coll. Interfac. Sci.* **27**, 305 (1968).

46. G. Rothenberger, J. Moser, M. Gratzel, N. Serpone, and D. K. Sharma, *J. Am. Chem. Soc.* **107**, 8054 (1985).

47. C. K. Gratzel, M. Jirousek, and M. Gratzel, *J. Mol. Catal.* **60**, 375 (1990).

48. D. T. Sawyer and M. J. Gibian, *Tetrahedron* **35**, 1471 (1979).

49. R. W. Matthews, *J. Phys. Chem.* **91**, 3328 (1987); R. W. Matthews, *Sol. Energy* **38**, 405 (1987); R. W. Matthews, *Aust. J. Chem.* **40**, 667 (1987).

50. M. Barbeni, C. Minero, E. Pelizzetti, E. Borgarello, and N. Serpone, *Chemosphere* **16**, 225 (1987).

51. C. Minero, C. Aliberti, E. Pelizzetti, R. Terzian, and N. Serpone, *Langmuir* **7**, 928 (1991).

52. J. Cunningham and S. Srijaranai, *J. Photochem. Photobiol. A: Chem.* **43**, 329 (1988).

53. J. R. Harbour, J. Tromp, and M. L. Hair, *Can. J. Chem.* **63**, 204 (1985).

54. O. I. Micic, Y. Zhang, K. R. Cromack, A. D. Trifunac, and M. C. Thurnauer, *J. Phys. Chem.* **97**, 7277 (1993).

55. R. W. Matthews, *J. Chem. Soc. Faraday Trans. 1* **80**, 457 (1984).

56. T. Sehlili, P. Boule, and J. Lemaire, *J. Photochem. Photobiol. A: Chem.* **50**, 117 (1989).

57. D. W. Bahnemann, in *Proceedings of the Symposium on Semiconductor Photoelectrochemistry*, C. Koval, ed., The Electrochemical Society, Inc., Pennington, N.J., 1991.

58. C. Richard and J. Lemaire, *J. Photochem. Photobiol. A: Chem.* **55**, 127 (1990).

59. T. Sehlili, P. Boule, and J. Lemaire, *J. Photochem. Photobiol. A: Chem.* **50**, 103 (1989).

60. M. Barbeni, E. Pramauro, E. Pelizzetti, E. Borgarello, M. Gratzel, and N. Serpone, *Nouv. J. Chim.* **8**, 547 (1984).

61. M. A. Fox, in Ref. 1, Chapt. 13, pp. 421–455.

62. M. A. Fox, R. B. Draper, M. Dulay, and K. O'Shea, in E. Pelizzetti and M. Schiavello, eds., *Photochemical Conversion and Storage of Solar Energy*, Kluwer Academic Publishers, Dordrecht, the Netherlands, 1991, pp. 323–336.

63. D. F. Ollis, E. Pelizzetti, and N. Serpone, in Ref. 1, p. 603.

64. H. Al-Ekabi and N. Serpone, *J. Phys. Chem.* **92**, 5726 (1988).

65. H. Al-Ekabi, N. Serpone, E. Pelizzetti, C. Minero, M. A. Fox, and R. B. Draper, *Langmuir* **5**, 250 (1989).

66. P. Pichat and J.-M. Herrmann, in Ref. 1, Chapt. 8, pp. 218–250.

67. R. W. Matthews, *Water Res.* **24**, 653 (1990).

68. C. S. Turchi and D. F. Ollis, *J. Catal.* **122**, 178 (1990).

69. R. W. Matthews and S. R. McEvoy, *J. Photochem. Photobiol. A: Chem.* **64**, 231 (1992).

70. M. W. Peterson, J. A. Turner, and A. J. Nozik, *J. Phys. Chem.* **95**, 221 (1991).

71. L. Sun, K.-M. Schindler, A. R. Hoy, and J. R. Bolton, *Proceedings of the Symposium on Environmental Aspects of Surface and Aquatic Photochemistry*, American Chemical Society Meeting, ACS, San Francisco, Calif., Apr. 5–10, 1992, p. 259.

72. C. Minero, F. Catozzo, and E. Pelizzetti, *Langmuir* **7**, 481 (1992).

73. G. L. Hug, *Optical Spectra of Nonmetallic Inorganic Transient Species in Aqueous Solution*, NSRDS-NBS 69, National Bureau of Standards, Washington, D.C., 1981.

74. D. Lawless, N. Serpone, and D. Meisel, *J. Phys. Chem.* **95**, 5166 (1991); D. Lawless, N. Serpone, and D. Meisel, in Ref. 57.

Nick Serpone
Concordia University

PHOTOCHROMISM.

See CHROMOGENIC MATERIALS.

PHOTOCONDUCTIVE MATERIALS.

See ELECTROPHOTOGRAPHY; PHOTOCONDUCTIVE POLYMERS; REPROGRAPHY.

PHOTOCONDUCTIVE POLYMERS

Polymers are, in general, insulators. Some conjugated polymers, such as polyacetylene, can be made into conductors by chemical doping (see ELECTRICALLY CONDUCTIVE POLYMERS). Strong oxidizing agents (eg, AsF₅) and reducing agents (eg, Na) have been used as dopants (1). The function of doping can also be performed by photons if photoactive molecules are present. Certain polymers are insulators in the dark but become conductive when irradiated by light. Poly(*N*-vinylcarbazole) [25067-59-8/ (PVK) was the first known photoconductive polymer (2). The discovery that PVK, when sensitized with dyes, exhibits photoconductivity high enough to be useful in electrophotography (qv) (2,3) has stimulated extensive research in this area.

Photoconductive polymers are widely used in the imaging industry as either photosensitive receptors or carrier (electron or hole) transporting materials in copy machines and laser printers. This is still the only area in which the photoelectronic properties of polymers are exploited on a large-scale industrial basis. It is also one electronic application where polymers are superior to inorganic semiconductors.

A good photoconductive polymer should exhibit the following properties. First, it has to be a good insulator in the dark and be capable of sustaining a high electric field. The superior dielectric strength of polymers, along with their good film-forming properties, are important reasons for their success in electrophotography. Secondly, when irradiated by light, the material has to generate carriers with high quantum efficiency. The charge generation efficiencies of most polymers are low and usually have to be enhanced by doping with electron donors or acceptors. Finally, the generated carriers have to move through the polymer film without being significantly trapped. Almost all known photoconductive polymers transport holes only.

It is useful to compare the sensitivity of electrophotography with other commonly employed imaging techniques (4) (Fig. 1). The sensitivity of electrophotography is not as high as that of silver halide but is better than many other techniques, such as photopolymerization. One of the main reasons for its high sensitivity is the ability of polymer to sustain high electrical fields, which in turn enhances both the quantum yield of charge separation and the carrier mobility. In the absence of an electric field, polymers usually have low charge generation efficiency and carrier mobility.

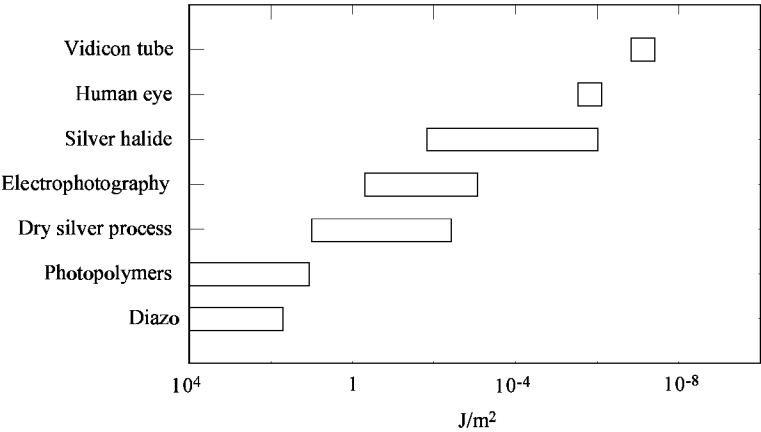


Fig. 1. Comparison of sensitivities of different imaging techniques (4).

There are many excellent reviews of photoconductive polymers (4–10). This article emphasizes results obtained after 1980, up to early 1994.

Classification of Materials

Photoconductive polymers can be conveniently classified into five categories based on their structures and modes of photoconduction.

Polymers With Pendent Groups. The classical photoconductive polymer, poly(*N*-vinylcarbazole) (PVK), belongs to this category. In this class of material, the polymer backbone does not participate in carrier transport directly. Instead, the carriers move by hopping along the electroactive pendent groups such as carbazole. These pendent groups are covalently attached to the polymers. With electron donors as the pendent group, the polymer conducts holes; with electron acceptors as the pendent group, the polymer conducts electrons. In general, only hole-conducting polymers have been successfully made with good enough properties for practical applications. The hole mobilities of this class of polymers are low, around 10⁻⁶–10⁻⁷ cm² Vs. Their intrinsic charge generation efficiency and spectral sensitivity range are also limited. Sensitizers such as dyes (2), 2,4,7-trinitro-9-fluorenone (11), and fullerenes (12,13) enhance the charge generation efficiency and extend the spectral range. Most of the sensitizers are electron acceptors, which form charge-transfer complexes with the donor groups of the polymers. The excitation of the charge-transfer complexes then leads to the generation of electrons and holes. Molecular structures of selected examples of this class of polymers are shown in Figure 2.

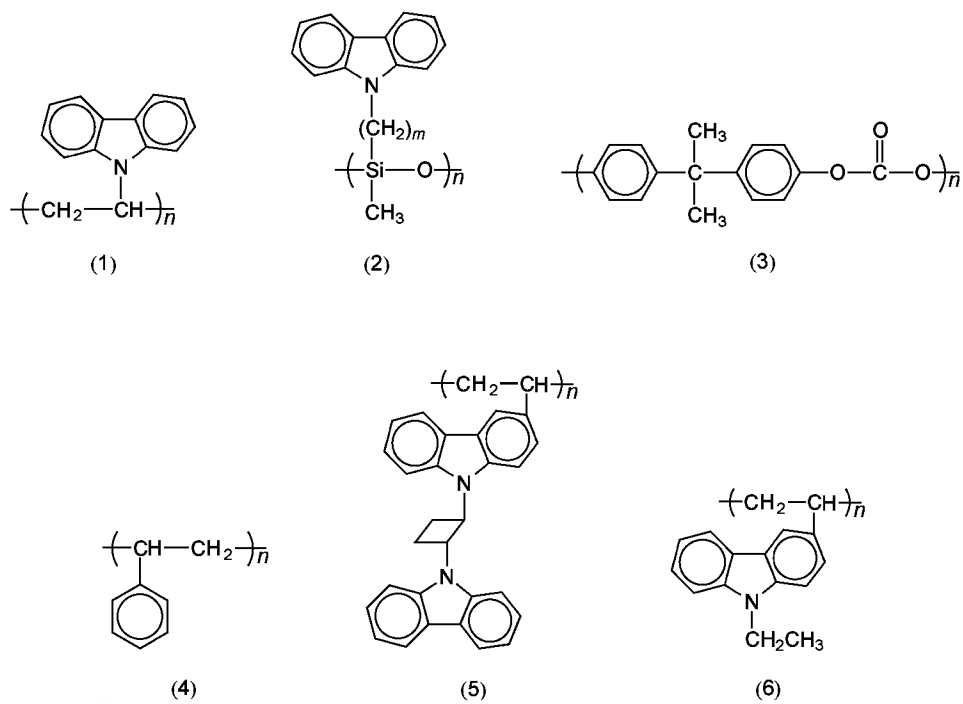


Fig. 2. Molecular structures of selected photoconductive polymers with pendent groups: (1) poly(*N*-vinylcarbazole) [25067-59-8] (PVK), (2) *N*-polysiloxane carbazole, (3) bisphenol A polycarbonate [24936-68-3], (4) polystyrene [9003-53-6], (5) polyvinyl(1,2-*trans*-bis(9*H*-carbazol-9-yl)cyclobutane) [80218-52-6] (PVCZB), and (6) poly(9-ethyl-3-vinylcarbazole) [25569-45-3] (P3VK).

Molecularly Doped Polymers. Many small molecules such as aromatic amines, eg, triphenylamine [603-34-9] (TPA), are excellent hole transport materials (see Table 1) (7). They can be used alone as the hole transporting layer in a device if they can be deposited as amorphous thin films. In general, however, it is more advantageous to mix them with polymers with high mechanical strength and good film-forming properties. A high concentration of hole transport molecules must be used so that the latter form an interconnecting conductive network. In this case, polymers merely act as the binders. They do not participate in carrier transport directly, but can affect the carrier mobility by modifying the trap depth and the distance between traps. The carrier mobility of this class of polymers is sensitive to the volume concentration of hole transport molecules present. Usually, the higher the concentration, the larger the hole mobility.

This class of polymeric photoconductors is distinguished from the pendent group-containing polymers, eg, PVK, by the fact that the active hole transport groups are not covalently bonded to the polymer backbone, but are merely dissolved in the polymer. This provides great flexibility for sample preparation. Different polymer matrices with different hole-transporting molecules can be combined without the need for difficult chemical synthesis. Two commonly used polymers are polycarbonate and polystyrene (Fig. 2). The molecular structures of a number of selected hole transporting molecules are shown in Figure 3.

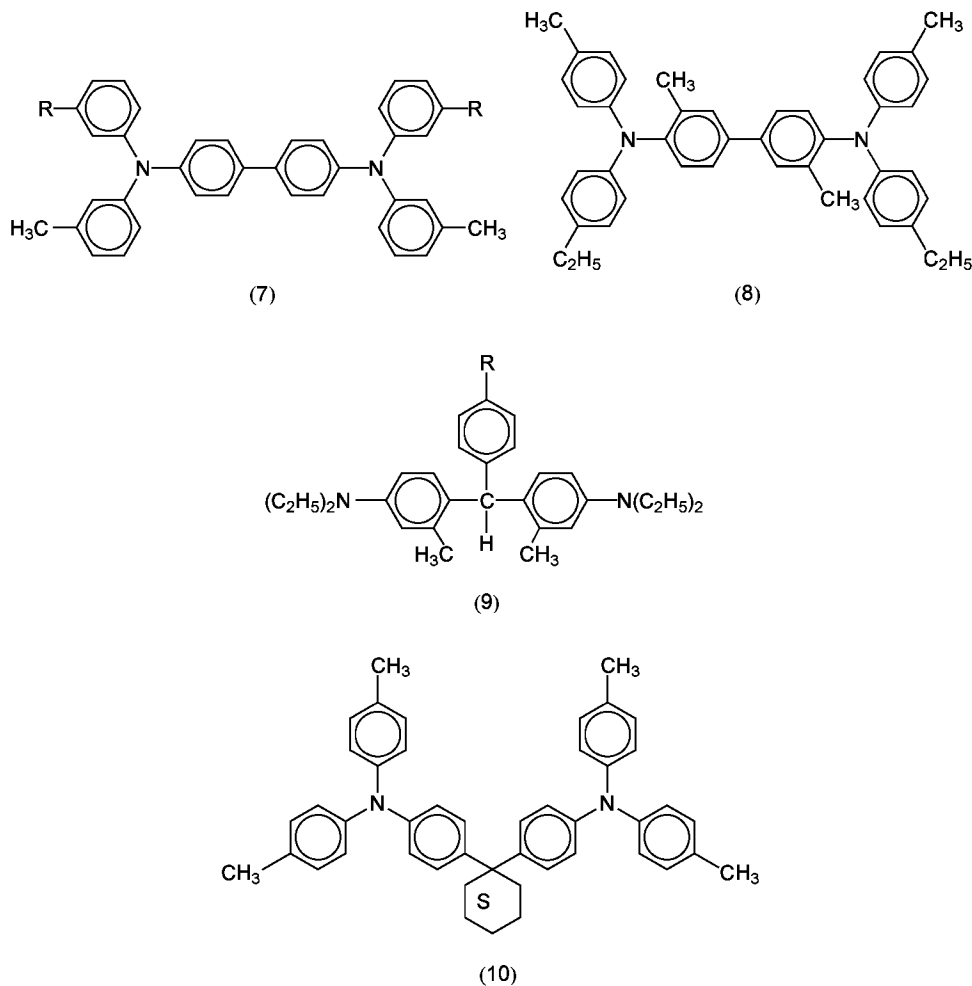


Fig. 3a. Molecular structures of selected hole-transport molecules. Biphenyls: (7a) (R = H)

N,N'-diphenyl-*N,N'*-bis(3-methylphenyl)-[1,1'-biphenyl]-4,4'-diamine [65181-78-4] (TPD); **(7b)** (R = CH₃) *N,N,N',N'*-tetrakis(4-methylphenyl)-(1,1'-biphenyl)-4,4'-diamine [106614-54-4] (TTB); **(8)** *N,N'*-bis(4-methylphenyl)-*N,N'*-bis(4-ethylphenyl)-[1,1'-(3,3'-dimethyl)biphenyl]-4,4'-diamine [115310-63-9] (ETPD). Triphenylmethanes (TPM) **(9)**, where R may = -N(C₂H₅)₂, -H, -OCH₃, -OH, -Br, -CN, -NO₂, and -CH₃; **(9)** (R = CH₃) is bis[4-(*N,N*-diethylamino)-2-methylphenyl] (4-methylphenyl)methane [70895-80-6] (MPMP); **(10)** 1,1-bis[(di-4-tolylamino) phenyl]cyclohexane [58473-78-2] (TAPC); **(11)** tetrakis-(3-methylphenyl)-*N,N,N',N'*-2,5-phenylenediamine [124591-08-8] (PDA); **(12)** α-phenyl-4-*N,N*-diphenylaminostyrene [89114-90-9] (TPS); **(13)** *p*-(diethylamino)benzaldehyde diphenylhydrazone [68189-23-1] (DEH); **(14)** 1-phenyl-3-[*p*-(diethylamino)styryl]5-[*p*-(diethylamino)phenyl] pyrazoline [57609-72-0] (PPR or DEASP); **(15)** 1,2-*trans*-bis(9*H*-carbazol-9-yl)cyclobutane [1484-96-4] (DCZB); and **(16)** 5'-[4-[bis(4ethylphenyl)amino]phenyl]-*N,N,N',N'*-tetrakis(4-ethylphenyl)-1,1':3',1''-terphenyl-4,4'' diamine (*p*-pEFTP).

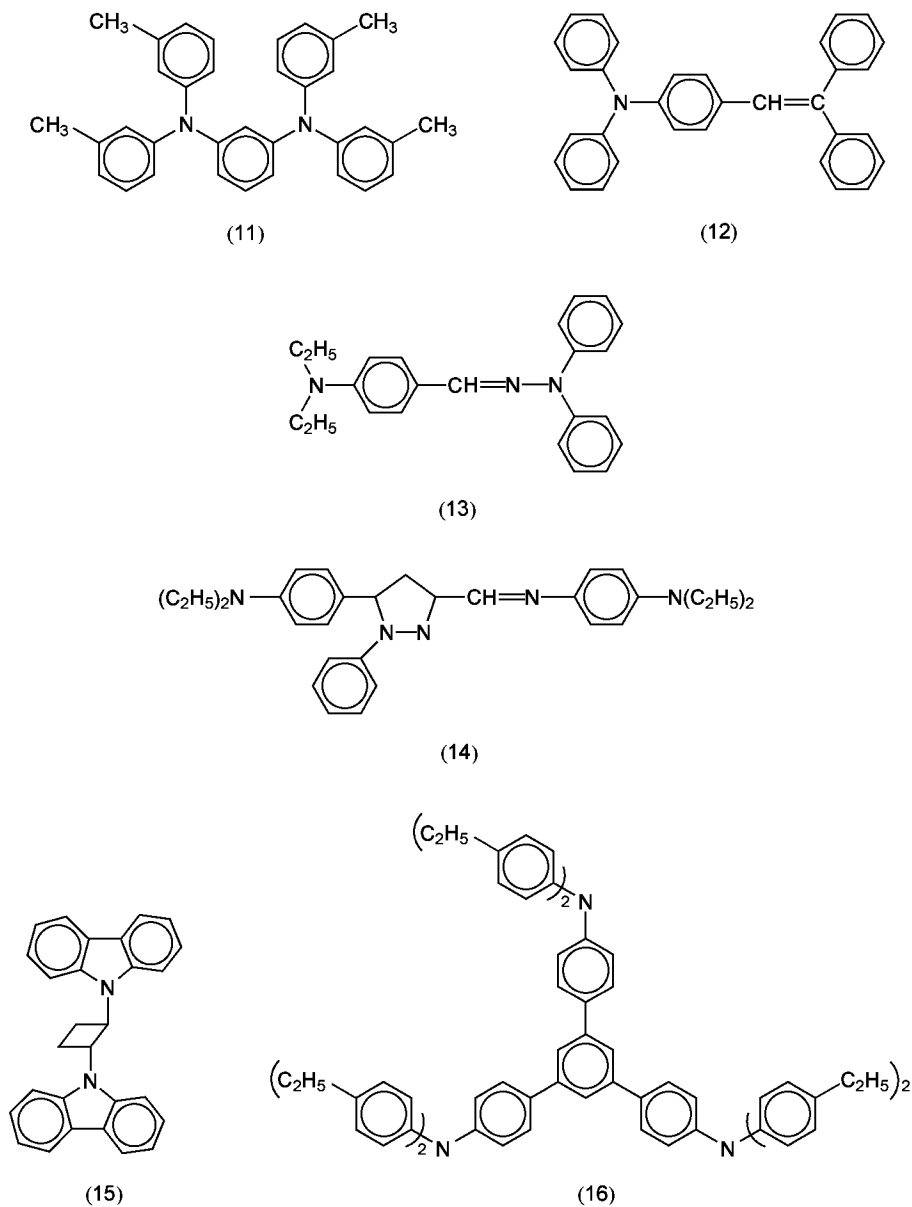
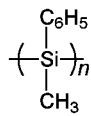


Fig. 3b. Continued

Backbone Conjugated Polymers. Polysilanes, (RR'Si)_n, are a unique class of polymers with the backbone consisting entirely of tetrahedrally coordinated silicon atoms. Extensive delocalization of σ-electrons takes place along the silicon chain, giving rise to many interesting electronic properties (14,15). Because of this σ-conjugation, carrier transport along the silicon backbone is very efficient. The hole mobility of polysilanes, ca 10⁻⁴ cm² Vs (16–19) is among the highest observed for polymers. Because the hole transport is through the σ-conjugated Si backbone, the hole mobility is insensitive to the substituent on the backbone. The hole mobilities of (phenylmethyl)polysilane (PMPS), poly(*n*-dodecylmethylsilane), poly(*n*-propylmethylsilane), and poly(methylcyclohexylsilane) are essentially the same (16,17). The charge generation efficiency and the spectral sensitivity range of polysilanes are, however, limited. Both can be enhanced by doping with sensitizers such as fullerenes (13).

Other polymers in this category include σ-conjugated polygermylenes (20) and π-conjugated polyacetylene, polythiophene, and poly(*p*-phenylenevinylene). The photoconductivity of many π-conjugated polymers can be enhanced by doping with fullerenes (21).

Liquid Crystalline Systems. Conventional photoconductive polymers are amorphous or systems with low order. In the case of PVK, the hole moves by hopping between the pendent carbazole groups. The hole mobilities are usually low, ~10⁻⁶ cm² Vs, due to a trap-dominated hopping transport (6–8). One approach to enhancing the hole mobility is to use conjugated polymers such as (phenylmethyl)polysilane (PMPS) (9–12). Another approach is the use of liquid crystalline systems where, in principle, transport can occur between ordered mesogenic groups (22–24).



The number of examples of liquid crystalline systems is limited. A simple discotic system, hexapentyloxytriphenylene **(17)** (Fig. 4), has been studied for its hole mobility (24). These molecules show a crystalline to mesophase transition at 69°C and a mesophase to isotropic phase transition at 122°C (25).

In the mesophase, the molecules exist in a discotic hexagonal columnar ordered structure, schematically shown in Figure 4.

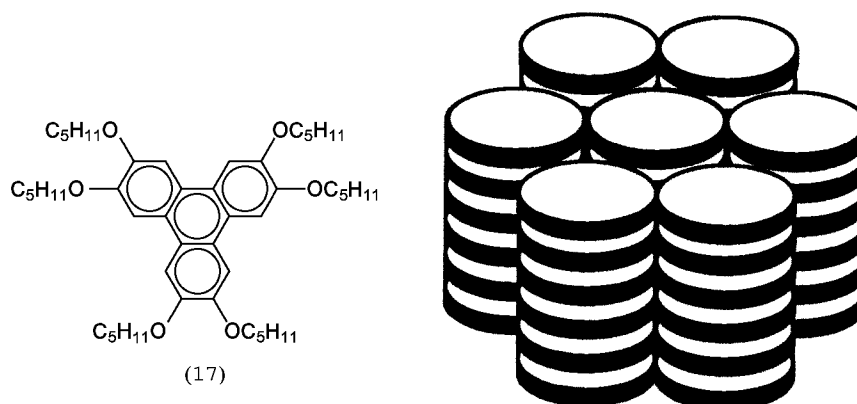


Fig. 4. Hexapentyloxytriphenylene (17) and a schematic view of its columnar mesophase (24).

The ordered columnar arrangement of the hexapentyloxytriphenylene molecules provides good overlap of the *p*-electrons of the triphenylene moieties along the director axis. This results in efficient hole transport in the mesophase. The hole photocurrent shows nondispersive transport with a high mobility up to $1 \times 10^{-3} \text{ cm}^2 \text{ Vs}^{-1}$ (24).

Nanoclusters/Polymer Composites. The principle for developing a new class of photoconductive materials, consisting of charge-transporting polymers such as PVK doped with semiconductor nanoclusters, sometimes called nanoparticles, Q-particles, or quantum dots, has been demonstrated (26,27).

The foundation for this new class of material is based on the ability to synthesize small semiconductor particles, typically in the nanometer-size regime (28–30). The structures of these semiconductor nanoclusters are usually the same as those of the bulk crystals, yet their properties are remarkably different. The electronic properties of these clusters depend on the cluster size, a phenomenon commonly referred to as the quantum size effect (28,29). It is manifested as a blue-shift in the exciton energy and enhancement in the volume-normalized oscillator strength as the cluster size decreases. An exciton is an electron–hole pair bound by Coulomb interaction. With the proper surface-capping agents, clusters of varying sizes can be isolated as powders and redissolved into various organic solvents in the same manner as molecules. By co-dissolving these clusters with the polymer, a thin film of nanocluster-doped polymer can be easily made by spin-coating. Alternatively, semiconductor nanoclusters can be directly synthesized in the polymer film (26–30).

So far polymers such as PVK, polysilane, and amine-doped polycarbonate have been used as the charge-transporting matrices (26,27). A wide variety of semiconductor nanoclusters have been synthesized within these polymers (26,27). Many narrow gap and *ir*-sensitive semiconductors such as InAs normally cannot be made into high field, room temperature photoconductors for electrophotography purposes. Other than the typical difficulty of growing good quality large area thin film, the main problem is the dark decay owing to thermal excitation of carriers. By dispersing nanometer-sized InAs in charge-transporting polymers, the charge-generation efficiency of InAs is retained, but the dark decay problem is removed. An additional benefit is the ease of thin-film preparation with polymers.

This new class of photoconductive nanocluster–polymer composites have not been extensively characterized and much work remains to be done. For example, the doping of semiconductor nanoclusters was shown to enhance the charge-generation efficiency of the polymer, but the effects on the transport properties are not yet known. Inorganic semiconductors such as CdS, Si, and Se have much higher ($>10^{-1} \text{ cm}^2 \text{ Vs}^{-1}$) carrier mobilities than organic polymers. Many of them are also electron-transport material, unlike polymers, which are usually hole-transport material. It is therefore of interest to examine the transport properties of composites containing small inorganic clusters embedded in polymers.

Charge Transport

For imaging applications such as electrophotography, the speed with which the carriers (electrons or holes) move through the photoconductor is less critical than for those applications involving serial processing such as photodetectors, where nanosecond or picosecond time resolution is often required. For electrophotography, typically the carriers need to move through a ca 10- μm film in milliseconds or tens of milliseconds. This requires the polymer to have a mobility of $>10^{-6} \text{ cm}^2 \text{ Vs}^{-1}$, which is easily achievable. More importantly, carriers have to move through the film without being trapped. It is quite remarkable that modern photoconductive polymers have been developed to the extent that polymer thin film without permanent traps can be easily fabricated. This is not the case for crystalline inorganic semiconductors, where carrier transport is extremely sensitive to the presence of impurities and fabrication of good quality thin film requires elaborate procedures.

The carrier mobility is usually measured by a time-of-flight method (31–33). The photoconductive film is sandwiched between two transparent electrodes, typically indium tin oxide and gold. The optical density of the film at the laser wavelength has to be high enough that when a low intensity, short-duration laser pulse irradiates the sample, the electron–hole pairs are created near the surface of the film. Depending on the polarity of the electric field applied on the electrodes, either electrons or holes traverse the bulk of the film. This gives rise to a displacement current that is detected by the external circuit, as shown in Figure 5a. In the ideal case (ie, in nondispersive transport), the current stays constant and falls off to zero at time τ_d when the charge carriers arrive at the other side of the film. Usually the fall-off near τ_d is smeared out due to spreading of the charge carriers packet (Fig. 5a). The carrier mobility, μ , is then determined from the equation: $\mu = l / (\tau_d E)$, where *l* is the film thickness and *E* is the applied field.

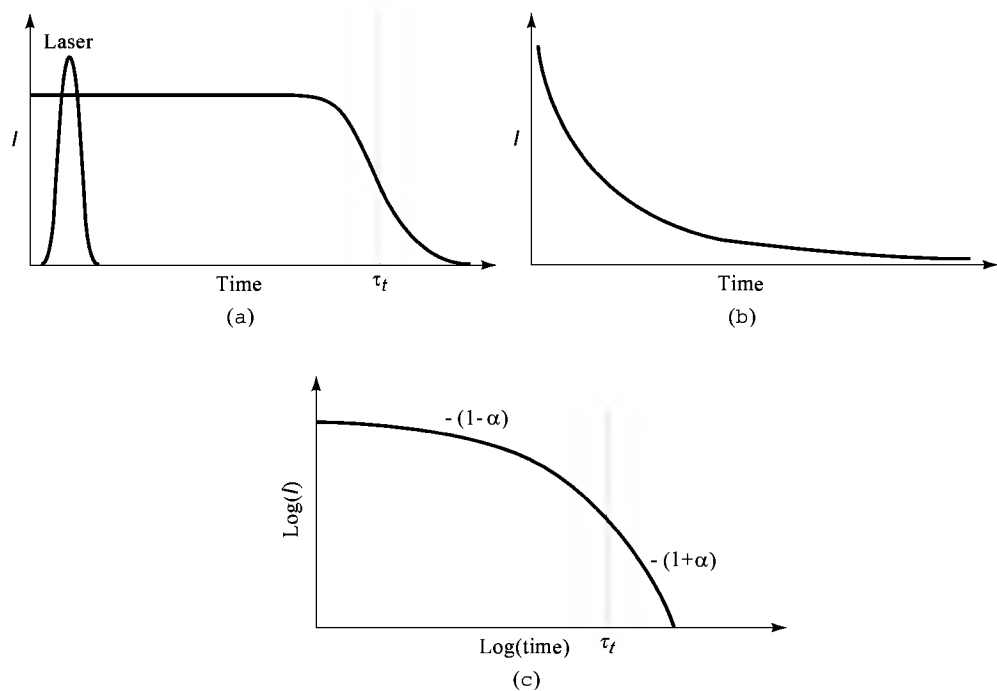


Fig. 5. Various current transients obtained by a time-of-flight method: (a) nondispersive transport; (b) dispersive transport; and (c) analysis of disperse current transient.

Carrier transport in polymers is characterized by a succession of hops from site to site. The distances between various neighboring sites and the energetics of each site are different from one another. These distributions (dispersions) in energy and distance cause different hopping rates between different sites. This is called dispersive transport, which gives a transient current that deviates from the ideal shape shown in Figure 5a. In the extreme case, the current looks like that shown in Figure 5b, without any discernible break at transit time τ_t . One of the central issues in photoconductive polymer research has been to develop a theoretical framework for understanding hopping transport in polymers.

Scher-Montroll Model. In an amorphous polymer there is a distribution in the separation distances between nearest-neighbor hopping sites as well as a distribution in the energy barriers between these sites. Both distributions contribute to a strong variation in hopping times. In the case of an extremely large hopping time dispersion, the transient current displays a featureless long tail, as shown in Figure 5b. Such dispersive charge transport was addressed in pioneering work done during the 1970s (34,35). It was proposed that the distribution of the hopping times, $\Psi(t)$, has the following form:

$$\Psi(t)_{t \rightarrow \infty} \sim t^{-(1+\alpha)} \quad 0 < \alpha < 1 \tag{1}$$

The transient current, derivable from equation 1, is given in equations 2 and 3 where τ_t is the transit time and I is the absorbed photon flux. The parameter α can be further derived as equation 4 (4), where T is the absolute temperature and T_0 is the distribution width (in units of kT) of a series of exponential traps. In this context, the carrier mobility is governed by trapping and detrapping processes at these sites.

$$I(t) \sim t^{-(1-\alpha)} \quad \text{for } t < \tau_t \tag{2}$$

$$I(t) \sim t^{-(1+\alpha)} \quad \text{for } t > \tau_t \tag{3}$$

$$\alpha = \frac{T}{T_0} \tag{4}$$

According to the Scher-Montroll model, the dispersive current transient (Fig. 5b) can be analyzed in a double-log plot of $\log(I)$ vs $\log(t)$. The slope should be $-(1-\alpha)$ for $t < \tau_t$ and $-(1+\alpha)$ for $t > \tau_t$ with a sum of the two slopes equal to 2, as shown in Figure 5c. For many years the Scher-Montroll model has been the standard model to use in analyzing dispersive charge transport in polymers.

Monte Carlo Simulation. Studies using the Monte-Carlo simulation technique (7,36,37) have shown great success in describing the charge transport properties of polymers. In this model, charge transport occurs by hopping through a manifold of localized states with both energy and positional disorder. For energetic disorder, the energies, E , of the localized states are assumed to have a Gaussian distribution, $(2\pi\sigma^2)^{-1/2} \exp(-E^2/2\sigma^2)$, where σ is the standard deviation (distribution width). The positional disorder is due to either the fluctuation in the intersite distances or the variation of the mutual orientation, or both. The extent of the positional disorder is simulated by allowing the wavefunction overlap parameter between the two sites to fluctuate in a random manner. No analytical solution can be found for hopping transport based on this model. An analytical theory based on an effective medium approach is available (38) but it has limited applicability.

With the Monte Carlo method, the sample is taken to be a cubic lattice consisting of $70 \times 70 \times 70$ sites with intersite distance of 0.6 nm. By applying a periodic boundary condition, an effective sample size up to 8000 sites (equivalent to 4.8- μm long) can be generated in the field direction (37,39). Carrier transport is simulated by a random walk in the test system under the action of a bias field. The simulation results successfully explain many of the experimental findings, notably the field and temperature dependence of hole mobilities (37,39).

Figure 6 shows the field dependence of hole mobility for TAPC-doped bisphenol A polycarbonate at various temperatures (37). The mobilities decrease with increasing field at low fields. At high fields, a $\log \mu \propto E^{1/2}$ relationship is observed. The experimental results can be reproduced by Monte Carlo simulation, shown by solid lines in Figure 6. The model predicts that the high field mobility follows the following equation (37) where $\hat{\sigma} = \sigma/kT$ (σ is the width of the Gaussian distribution density of states), Σ is a parameter that characterizes the degree of positional disorder, E is the electric field, μ_0 is a prefactor mobility, and C is an empirical constant given as $2.9 \times 10^{-4} (\text{cm V}^{-1})^{1/2}$.

$$\mu(\hat{\sigma}, \Sigma, E) = \mu_0 \exp \left[\left(\frac{2\hat{\sigma}}{3} \right)^2 \right] \exp \left[C(\hat{\sigma}^2 - \Sigma^2) E^{1/2} \right] \tag{5}$$

This equation has been used to analyze many experimental mobility data successfully, and the resultant fitting parameters are tabulated in Table 1. The

value of the prefactor mobility depends on the form of the temperature dependence used. Use of a T^{-1} dependence instead of the T^{-2} dependence of equation 5 results in values that are orders-of-magnitude higher.

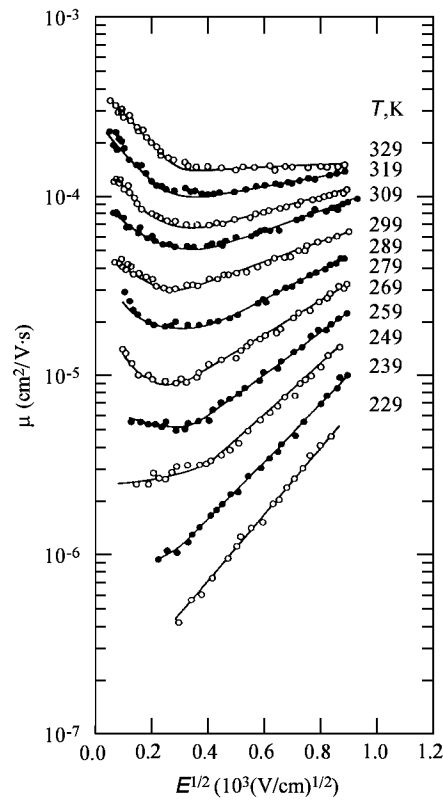


Fig. 6. The logarithm of the mobility vs $E^{1/2}$, parametric in temperature, for TAPC-doped polycarbonate (37).

Based on the Monte Carlo simulations, it is seen that the presence of positional disorder causes the mobility to decrease with increasing field at low fields (37). This is the case because the introduction of positional disorder into the system provides the carrier with energetically more favorable routes, which occasionally are against the field direction. These detour routes are most efficient at low fields, but are eliminated at high fields. This rationalizes the decrease of hole mobilities with increasing field.

One potential weakness of the disorder formalism is the neglect of the polaronic effect. When an electron is transferred from the neutral form of the transport molecule to the ionic form, there is a distortion in the molecular structure, ie, the formation of a polaron, in physicists' language. The polaron model considers this intramolecular deformation energy to be more important than the disorder energy (40–43). In general, such deformation must occur during charge transport and should be considered in the next level of the disorder formalism. Guidelines for estimating the polaronic effect have been discussed and it is believed that this effect usually is small (7).

Experimental Hole Mobilities. The experimental values of hole mobilities in polymers are tabulated in Tables 1 and 2. The hole mobility is field dependent. Whenever the experimental data have been fitted with equation 5, the parameters μ_0 , σ , and σ , which give a complete description of the field dependence of the hole mobility, are listed (Table 2). Otherwise, hole mobilities at selected fields are listed. All acronyms are defined in Figures 2 and 3.

The values of the hole mobilities for polymers span a large range, from ca 10^{-10} cm² Vs to ca 10^{-1} cm² Vs. Polymers with pendent groups such as PVK tend to have the smallest mobilities (Table 1, entries 1 to 13). The mobility depends on the average separation distance between the pendent groups (Table 1, entries 6–8) (47). Polymers with conjugated backbones (Table 1, entries 14–23) and molecularly doped polymers (Table 1, entries 24–50 and all Table 2 entries) generally have higher mobilities. As a class, the transport properties of molecularly doped polymers have been studied more systematically because of the flexibility in sample fabrication. Several interesting effects and trends have been observed.

Table 1. Hole Mobilities of Selected Molecular and Polymeric Materials^a

Entry	Materials	Hole mobility, cm ² Vs	Reference
1	PVK	10^{-6} at 5×10^5 V cm; 10^{-7} at 5×10^4 V cm	44
2	10% TNF PVK	3×10^{-6} at 1×10^6 V cm; 10^{-7} at 1×10^5 V cm	11
3	1% C-60 PVK	2×10^{-7} at 3×10^5 V cm; 10^{-6} at 7×10^5 V cm	45
4	P3VK	1.9×10^{-7} at 4×10^5 V cm	46
5	PVDCZB	3.2×10^{-6} at 4×10^5 V cm	46
6	N-polysiloxanecarbazole, <i>m</i> = 3	ca 10^{-7} at 2.5×10^5 V cm and 263 K	47
7	N-polysiloxanecarbazole, <i>m</i> = 5	ca 2×10^{-8} at 2.5×10^5 V cm and 263 K	47
8	N-polysiloxanecarbazole, <i>m</i> = 6	ca 10^{-8} at 2.5×10^5 V cm and 263 K	47
9	0.01% TPD PVK	ca 3×10^{-7} at 3×10^5 V cm	48,49
10	1% TPD PVK	ca 3×10^{-9} at 3×10^5 V cm	48,49
11	60% TPD PVK	ca 10^{-4} at 3×10^5 V cm	48,49
12	PTPB	10^{-5} at 3×10^5 V cm	50
13	poly(hydroxyamino ester) (PHA)	2×10^{-6} at 8×10^5 V cm	51
14	(phenylmethyl)polysilane (PMPS)	10^{-4} at 2×10^5 V cm; 3×10^{-4} at 1×10^6 V cm	16,17
15	1% C-70 PMPS	9×10^{-5} at 3×10^4 V cm; 1.4×10^{-3} at 4×10^5 V cm	45
16	1% TPD PMPS	ca 10^{-4} at 3×10^5 V cm	48
17	1% DEH PMPS	ca 10^{-6} at 3×10^5 V cm	48
18	1% PPR PMPS	ca 5×10^{-8} at 3×10^5 V cm	48
19	50% TPS PMPS	10^{-3} at 2×10^5 V cm	52
20	50% PDA PMPS	10^{-3} at 2×10^5 V cm	52
21	60% TPD PMPS	8×10^{-4} at 2.5×10^5 V cm	48
22	75% ETPD PMPS	ca 10^{-1} at 2.5×10^5 V cm	48

23	PDBG	ca 10 ⁻⁴	20,53
24	10 ⁰ ◦ TPD polystyrene	ca 10 ⁻⁸ at 10 ⁵ V cm	54
25	20 ⁰ ◦ TPD polystyrene	ca 10 ⁻⁶ at 10 ⁵ V cm	54
26	50 ⁰ ◦ TPD polystyrene	ca 10 ⁻⁴ at 10 ⁵ V cm	54
27	75 ⁰ ◦ TPD polystyrene	ca 2 × 10 ⁻⁴ at 10 ⁵ V cm	54
28	10 ⁰ ◦ TPD polycarbonate	ca 10 ⁻¹⁰ at 10 ⁵ V cm	54
29	75 ⁰ ◦ TPD polycarbonate	ca 5 × 10 ⁻⁵ at 10 ⁵ V cm	54
30	10 ⁰ ◦ ETPD polystyrene	ca 10 ⁻⁷ at 10 ⁵ V cm	54
31	20 ⁰ ◦ ETPD polystyrene	ca 10 ⁻⁵ at 10 ⁵ V cm	54
32	50 ⁰ ◦ ETPD polystyrene	ca 5 × 10 ⁻⁴ at 10 ⁵ V cm	54
33	75 ⁰ ◦ ETPD polystyrene	ca 10 ⁻³ at 10 ⁵ V cm	54
34	10 ⁰ ◦ ETPD polycarbonate	ca 10 ⁻⁸ at 10 ⁵ V cm	54
35	75 ⁰ ◦ ETPD polycarbonate	ca 10 ⁻⁴ at 10 ⁵ V cm	54
36	70 ⁰ ◦ DEASP polystyrene	6 × 10 ⁻⁷ at 4 × 10 ⁴ V cm; 2 × 10 ⁻⁵ at 6 × 10 ⁵ V cm	43
37	10 ⁰ ◦ DEASP polystyrene	4 × 10 ⁻⁹ at 1 × 10 ⁶ V cm	43
38	70 ⁰ ◦ DEASP polycarbonate	6 × 10 ⁻⁷ at 4 × 10 ⁴ V cm; 2 × 10 ⁻⁵ at 6 × 10 ⁵ V cm	43
39	10 ⁰ ◦ DEASP polycarbonate	1.3 × 10 ⁻⁹ at 1 × 10 ⁶ V cm	43
40	50 ⁰ ◦ DEH polycarbonate	ca 1 × 10 ⁻⁷ at 6 × 10 ⁴ V cm; ca 1 × 10 ⁻⁶ at 1 × 10 ⁶ V cm	55
41	20 ⁰ ◦ TPA polycarbonate	10 ⁻⁶ at 5 × 10 ⁵ V cm	56,57
42	45 ⁰ ◦ TPA polycarbonate	10 ⁻⁵ at 5 × 10 ⁵ V cm	56,57
43	TPM, R = -N(C ₂ H ₅) ₂	8 × 10 ⁻⁴ at 5 × 10 ⁵ V cm	58
44	TPM, R = -H	1 × 10 ⁻⁵ at 5 × 10 ⁵ V cm	58
45	TPM, R = -OCH ₃	3 × 10 ⁻⁶ at 5 × 10 ⁵ V cm	58
46	TPM, R = -OH	7 × 10 ⁻⁷ at 5 × 10 ⁵ V cm	58
47	TPM, R = -Br	7 × 10 ⁻⁷ at 5 × 10 ⁵ V cm	58
48	TPM, R = -CN	5 × 10 ⁻⁸ at 5 × 10 ⁵ V cm	58
49	TPM, R = -NO ₂	2 × 10 ⁻⁸ at 5 × 10 ⁵ V cm	58
50	thiapyrylium dye aggregates in TPM polycarbonate	ca 10 ⁻⁸ at ca 10 ⁵ V cm	59
51	discotic mesophase of hexa-pentyloxytriphenylene (HPT)	10 ⁻³ from 1 × 10 ⁴ to 3 × 10 ⁵ V cm	24
52	DCZP poly(11-(4-cyano-4'-biphenyl)-1-undecanoyl acrylate	10 ⁻⁶ at 6 × 10 ⁵ V cm	60

^a All acronyms are defined in Figures 2 and 3.

In molecularly doped polymers, charge transport is carried out by the hole-transporting molecular dopants, usually aromatic amines. The polymer merely acts as a binder. The hole mobility is sensitive to the dopant concentrations. For example, the hole mobility of *N,N'*-diphenyl-*N,N'*-bis(3-methylphenyl)-[1,1'-biphenyl]-4,4'-diamine [65181-78-4] (TPD) (7a) in polystyrene is ca 10⁻⁸ cm² Vs at 10⁵ V cm with 10⁰ ◦ TPD concentration but increases to ca 2 × 10⁻⁴ cm² Vs with 75⁰ ◦ TPD concentration (entries 24–27 of Table 1) (54).

Although polymers do not participate in carrier transport directly, they can affect the carrier mobility by modifying the carrier trap depth and distribution. For example, the hole mobility of 1,1-bis[(di-4-tolylamino) phenyl]cyclohexane [58473-78-2] (TAPC) (10) doped polystyrene is about 100 times larger than that of TAPC-doped polycarbonate (Table 2, entries 2 and 3) (37). The σ-parameter, which measures the energetic disorder, decreases when polycarbonate is replaced by polystyrene. This change is attributed to the elimination of dipolar fields generated by the carbonyl groups of the polycarbonate (62). It was also observed that the μ₀ parameter increases by an order of magnitude in conjunction with a decrease in σ from polycarbonate to polystyrene (Table 2). This suggests that intermolecular electronic coupling between TAPC molecules is improved by the reduction of positional disorder in polystyrene (62). The same effect was also observed when copolymers of styrene and butyl acrylate were used as the matrices (63). It was found that the higher the styrene content, the smaller the σ-value and the higher the mobility (Table 2, entries 4–8). This is attributed to the increase of energetic disorder owing to the polar carbonyl groups in butyl acrylate (63). Studies using other hole-transporting molecules such as TPD and ETPD show the same striking effects on the hole mobility by the host polymers (Table 1, entries 24–35) (54).

Table 2. μ₀,σ, and Σ and Some Hole Mobilities, cm²/V s, of Selected Molecular and Polymeric Materials^a

Entry	Materials	μ ₀ , cm ² Vs	σ, eV	Σ	Ref.
1	1,1-bis[(di-4-tolylamino) phenyl] cyclohexane (TAPC)	0.13	0.068	1.0	61
2	75 ⁰ ◦ TAPC polystyrene ^b	0.23	0.077	2.14	62
3	75 ⁰ ◦ TAPC polycarbonate ^c	0.02	0.095	3.1	62
4	40 ⁰ ◦ TAPC styrene:butylacrylate copolymers, S:BA = 100:0	1.5 × 10 ⁻²	0.069	2.4	63
5	40 ⁰ ◦ TAPC styrene:butylacrylate copolymers, S:BA = 95:5	1.7 × 10 ⁻²	0.073	2.4	63
6	40 ⁰ ◦ TAPC styrene:butylacrylate copolymers, S:BA = 85:15	1.7 × 10 ⁻²	0.08	2.4	63
7	40 ⁰ ◦ TAPC styrene:butylacrylate copolymers, S:BA = 75:25	1.6 × 10 ⁻²	0.086	2.4	63
8	40 ⁰ ◦ TAPC styrene:butylacrylate copolymers, S:BA = 50:50	1.5 × 10 ⁻²	0.103	2.4	63
9	40 ⁰ ◦ TAPC poly(4- <i>tert</i> -butylstyrene)	1.5 × 10 ⁻²	0.077	2.4	64
10	40 ⁰ ◦ TAPC poly(4-chlorostyrene)	1.3 × 10 ⁻²	0.087	2.4	64
11	<i>N,N'</i> -diphenyl- <i>N,N'</i> -bis(3-methyl-phenyl)-[1,1'-biphenyl]-4,4'-diamine (TPD) ^d	0.035	0.074	1.2	61,65
12	1-phenyl-3-[<i>p</i> -(diethylamino)styryl]-5-[<i>p</i> -(diethylamino)phenyl] pyrazoline (DEASP or PPR)	0.006	0.103	1.4	61
13	<i>p</i> -(diethylamino)benz-aldehyde diphenylhy-drazone (DEH)	0.0013	0.1	2.0	61
14	<i>N,N,N',N'</i> -tetrakis(4-methyl-phenyl) (1,1'-biphenyl)-4,4'-diamine (TTB)	0.019	0.069	1.5	61
15	bis[4-(<i>N,N</i> -diethylamino)-2-methyl-phenyl] (4-methylphenyl)-methane (MPMP)	0.34	0.098	2.0	61
16	5'-[4-[bis(4-ethylphenyl)amino]-phenyl]- <i>N,N,N',N'</i> -tetrakis(4-ethylphenyl)-1,1':3',1''-terphenyl-4,4''-diamine (<i>p-p</i> EFTP)	0.069	0.068	1.0	66

^a The hole mobility is field-dependent and only selected low field and high field values are listed here. All data were measured at room temperature unless

otherwise noted. All acronyms are defined in Figures 2 and 3.

^b Hole mobilities, cm² Vs, 6 × 10^{−3} at 4 × 10⁴ V cm; 1 × 10^{−2} at 5 × 10⁵ V cm.

^c Hole mobilities, cm² Vs, 4 × 10^{−5} at 4 × 10⁴ V cm; 6 × 10^{−5} at 5 × 10⁵ V cm.

^d Hole mobility ca 10^{−3} at 10⁵ V cm.

The effects of the dipole moment on the hole mobility are clearly illustrated in a study using a series of hole-transporting amines with different dipole moments (61). The mobility data can be analyzed using equation 5. It was found that amines with a large dipole moment have large σ and lower mobility (Table 2, entries 1 and 11–14) (61). This again confirms the detrimental effects of polar groups on the carrier mobility in amorphous systems.

Interesting effects are observed when "transport-active" polymers are treated with various dopants. At low dopant concentrations, the hole mobility is expected to be the same as that of the host polymer with minimum perturbation from the dopant (eg, Table 1, entries 3, 9, and 15). However, even at low dopant concentrations, decreases in hole mobility have been observed when the oxidation potential of the dopant is much lower than that of the host. For example, the hole mobility of (phenylmethyl)polysilane [76188-55-1] (PMPS) scales qualitatively with the oxidation potential of the dopants TPD, *p*-(diethylamino)benzaldehyde diphenylhydrazone [68189-23-1] (DEH) (13), and 1-phenyl-3-[*p*-(diethylamino)styryl]-5-[*p*-(diethylamino)phenyl] pyrazoline [57609-72-0] (PPR) (14) (all 1° o doping) (48). A hole mobility of 10^{−4} cm² Vs exists for TPD-doped PMPS, while PPR-doped PMPS has a mobility of only 5 × 10^{−8} cm² Vs (Table 1, entries 16–18). The dopant PPR, having a low oxidation potential, acts as a deep trap and reduces the hole mobility.

At very high dopant concentrations, transport occurs directly between the dopant molecules. The polymer acts only as a binder in most cases. Taking TPD-doped PVK as an example, at low TPD concentrations the hole mobility first decreases from 3 × 10^{−7} cm² Vs to 10^{−9} cm² Vs with increasing TPD concentration, because TPD molecules act as hole traps (48,49). At higher TPD concentrations, new direct transport channels between the TPD molecules open up and the hole mobility increases to 10^{−4} cm² Vs for ca 60° o TPD doping (Table 1, entries 9–11) (48,49). In this case, there is no evidence for unusual interaction between TPD and PVK that affects the hole transport process.

However, certain synergistic effects have been observed recently in several systems when transport-active polymers are doped with high concentrations of hole-transporting molecules (48,52). For example, doping with high concentrations of α -phenyl-4-*N,N*-diphenylaminostyrene [89114-90-9] (TPS) (12) (52), tetrakis(3-methylphenyl)-*N,N,N',N'*-2,5-phenylenediamine [124591-08-8] (PDA) (11) (52), and TPD (7a) (48) in PMPS produces samples with hole mobilities of ca 10^{−3} cm² Vs (Table 1, entries 19–21), which are higher than in neat undoped PMPS and also higher than these dopants would show if they were dispersed in transport-inactive matrices at the same concentrations. In one instance, *N,N'*-bis(4-methylphenyl)-*N,N'*-bis(4-ethylphenyl)-[1,1'-(3,3'-dimethyl)biphenyl]-4,4'-diamine [115310-63-9] (ETPD) (8) -doped PMPS (Table 1, entry 22), hole mobility approaching 10^{−1} cm² Vs at 2.5 × 10⁵ V cm was observed (48). This is the highest recorded hole mobility for disordered organic systems. From this perceptive, it is very interesting to study the carrier mobility of polymers heavily doped with semiconductor nanoclusters.

Charge Generation

Another important property of a photoconductor is the efficiency with which it converts photons into electrons and holes. Most of the photoconductive polymers have low charge-generation efficiencies. This can be a result either of an intrinsically low charge-generation efficiency or of a low absorption coefficient in the interesting spectral region. To enhance the charge-generation efficiency, sensitizers have to be added. Photoexcitation of the sensitizer generates an excited state, which can be either a singlet, triplet, or charge-transfer state. The excited state can relax radiatively or nonradiatively, or undergo electron-transfer reaction. If it accepts an electron from the surrounding polymer, a hole is then generated in the polymer. This initially generated electron–hole pair may recombine to the neutral ground state, or separate under the electric field into free carriers for conduction.

The charge-generation efficiency of a photoconductor can be measured by the standard photoinduced discharge method (33,67,68). The sample film is deposited on an electrically grounded aluminum substrate and corona-charged positively or negatively in the dark. The voltage is detected by an electrostatic voltmeter. Absorption of light generates electrons and holes which migrate to the surface and discharge the voltage if the sample is photoconductive. For light of sufficiently low intensity absorbed within a small fraction of the film thickness (the so-called emission-limited condition), the charge-generation efficiency, ϕ , can be obtained from the initial discharge rate of the surface potential, $(dV/dt)_{t=0}$ (eq. 6) (33,67,68). Here ϵ is the dielectric constant, e the electronic charge, L the film thickness, and I the absorbed photon flux. This technique works best for materials with minimal deep carrier traps.

$$\phi = \frac{\epsilon}{4\pi e L I} \left(\frac{dV}{dt} \right)_{t=0}$$

(6)

The details of the charge-generation mechanism can be very complicated, requiring a complete understanding of the primary photophysical and photochemical processes. Systematic studies on the charge-generation mechanism are few (69–71). Furthermore, unlike charge transport where the theoretical framework for detailed understanding is available, there is no theoretical model that can adequately address charge recombination and charge separation under electric field quantitatively.

Onsager Model. The theory developed by Onsager (72) has been the standard model to use for analyzing the electric field dependence of the charge-generation efficiency. The model solves the diffusion equation of the relative motion of an electron–hole pair, bounded by their Coulomb interaction, under an electric field. The origin of the electron–hole pair and the pathway by which it is generated are not considered in this model. The model solves for the probability that the pair separates toward infinity with a given initial separation distance, r_0 . An important boundary condition and assumption for this model is that if the pair separation distance reaches zero, the pair recombines immediately. With this assumption, the charge-generation efficiency, $\phi(r_0, E)$, in the presence of an electric field, E , is given by the following (72,73):

$$\phi(r_0, E) = \phi_0 \left\{ 1 - (2\xi)^{-1} \sum_{j=0}^{\infty} A_j(\eta) A_j(2\xi) \right\}$$

(7)

where

$$\eta = \frac{e^2}{\epsilon k T r_0}$$

(8)

$$2\xi = \frac{e E r_0}{k T}$$

(9)

$$A_{j+1}(\eta) = A_j(\eta) \frac{\eta^{j+1} e^{-\eta}}{(j+1)!}$$

(10)

$$A_0(\eta) = 1 - e^{-\eta}$$

(11)

Here ϕ_0 is the quantum yield of the initially generated electron–hole pair. The two parameters, r_0 and ϕ_0 , characterize quantitatively the charge generation efficiency of a photoconductor under applied field. For example, a large r_0 value indicates the photoconductor has a large low field charge-generation efficiency while the ϕ_0 value represents the ultimate charge-generation efficiency achievable at high field.

The field dependences of the charge-generation efficiency of many polymeric photoconductors have been analyzed in terms of the Onsager model and found to be in satisfactory agreement. For example, Figure 7 shows the field dependence of the charge-generation efficiencies of fullerene-doped PVK and PMPS. The corresponding fits with the Onsager model yield $r_0 = 1.9$ nm and $\phi_0 = 0.9$ for fullerene-doped PVK; $r_0 = 2.7$ nm and $\phi_0 = 0.85$ for fullerene-doped PMPS (12,13). This indicates that, in order to explain the field-dependence of the charge-generation efficiency, a fairly large initial electron–hole separation distance of 1.9–2.7 nm must be assumed. For polymeric photoconductors, this finding is quite common. Fitting with the Onsager model always results in large initial electron–hole separation distances with these materials (Table 3). The problem is that such a large initial electron–hole separation distance is improbable in a molecular system. In fact, in many cases it is concluded that charge-transfer state is responsible for the charge generation, which should lead to an initial electron–hole separation distance of only a few tenths of a nanometer. This paradox indicates the inadequacy of the Onsager model and is still an unresolved issue.

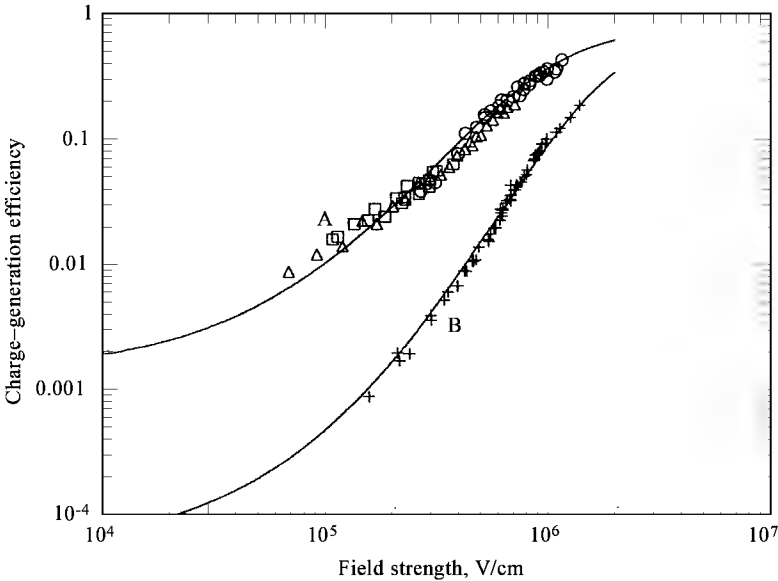


Fig. 7. The field-dependence of the charge-generation efficiency of a 2.0- μm thick ($\circ$), a 1.1- μm thick ($\square$), and 1.8- μm thick ($\triangle$) fullerene-PMPS film obtained with positive charging and 340-nm irradiation (A). The solid lines are calculated from the Onsager model. The best-fit curve is obtained with $r_0 = 2.7$ nm and $\phi_0 = 0.85$. Also plotted is the charge-generation efficiency of a fullerene-PVK film (+) obtained with positive charging and 340-nm irradiation (B). The solid lines are calculated from the Onsager model. The best-fit curve is obtained with $r_0 = 1.9$ nm and $\phi_0 = 0.9$ (13).

Modified Onsager Models. The boundary condition used in the Onsager model corresponds to the assumption that when the separation distance of an electron–hole pair reaches zero, it recombines with an infinitely fast rate. This assumption is unrealistic in real systems. For an electron–hole pair to recombine to the ground state, several electron volts of energy have to be disposed of, usually by being dumped into the vibrational modes of the system. This can be a slow (eg, nanoseconds) process because of the large amount of energy involved. Next, the creation and recombination rate may depend on the field, separation distance, and energetics. None of these are considered by the Onsager model.

The need for a finite recombination rate has been recognized (81,82). One suggested model identifies the geminate electron–hole pair with the excited charge-transfer state which typically has a lifetime on the scale of nanoseconds (81). In another model, the escape probability of an electron–hole pair is solved for using a different boundary condition than the original Onsager model (82). The boundary condition corresponds to a finite surface recombination velocity on a partly absorbing sphere of finite radius. Indeed, by assuming a slow recombination rate (taken to be the lifetime of either the charge-transfer state or the singlet state), the field dependence data can be fitted by these modified Onsager models with a small initial electron–hole separation distance (81,82). However, in these models there is no explanation for the origin of the slow recombination rate and the electric field dependence is not considered. Furthermore, there is no theory for either the creation or the escape rate of the electron–hole pair. Rather, a somewhat arbitrary form for the escape rate is assumed (82).

In principle, Marcus electron–transfer theory can be used to describe the creation and recombination of an electron–hole pair. The Marcus electrontransfer theory can qualitatively explain the origin of the slow recombination rate for a fullerene-doped PVK photoconductor (29). A careful examination of the energetics of the electron-transfer reaction between fullerene and carbazole reveals that the forward electron transfer is exothermic by 0.37 eV, corresponding to the maximum rate region in the Marcus electron-transfer theory (83). However, the backward electron transfer (recombination) is exothermic by 1.52 eV, which falls in the so-called Marcus inverted region (83). In the inverted region, the recombination rate is slower by many orders-of-magnitude. This provides the basic reason for the efficient charge separation in fullerene-doped polymers under applied field. For quantitative comparison with the experimental data, an Onsager-type model incorporating field-dependent Marcus electron-transfer theory needs to be developed.

Table 3. Charge-Generation Efficiency of Selected Polymeric Photoconductors^a

Polymers	r_0 , nm	ϕ_0	Reference
PVK	2.25	0.14 at 345 nm	74
0.1% trichloroacetic acid-PVK	3.0	0.11 at 345 nm	74
6 mol % 2,4,7-trinitro-9-fluorenone-PVK	2.5	0.23 at 550 nm	75
50 mol % 2,4,7-trinitro-9-fluorenone-PVK	3.5	0.23 at 550 nm	75
16% α -form metal-free phthalocyanine-PVK		0.4 at 620 nm (1.4×10^6 V/cm)	76
2.7% fullerene-PVK	1.9	0.9 at 340 nm	12
1.6% fullerene-PMPS	2.7	0.85 at 340 nm	13
poly(hydroxyamino ester) (PHA)	3.5	0.02 at 380 nm	51
32% CBr ₄ -PHA	2.4	0.04 at 380 nm	51
32% CBr ₄ -PHA + Michler's hydrol blue photolysis products	2.2	1.0 at 630 nm	51

thiapyrylium dye aggregates in TPM polycarbonate	4.4	0.58 at 680 nm	77
DPBDK TPA polycarbonate	3.3	0.53 at 300 nm	78
1 vol % 1.6-nm CdS clusters PVK	2.6	0.16 at 340 nm	26
BZ perylene		0.1 (8×10^5 V cm)	84
BZ perylene TPD bilayer		0.1 (3×10^4 V cm) 0.6 (8×10^5 V cm)	79
TPATA		0.06 (2×10^5 V cm)	80
TPATA amine bilayer		0.01 (5×10^4 V cm) 0.8 (8×10^5 V cm)	80

^a The initial electron–hole separation distance, r_0 , and the quantum yield, ϕ_0 , are derived by fitting with the Onsager model. When the initial quantum yield, ϕ_0 , is not known, the quantum yield at a specified field (in parentheses) is listed. All acronyms are defined in Figures 2 and 3.

Experimental Values of Charge-Generation Efficiencies. In this section the charge-generation efficiencies of many polymeric photoconductors are compared (Table 3). When the experimental data has been fitted to the Onsager model, the initial electron–hole separation distance, r_0 , and the initial quantum yield, ϕ_0 , are listed. Otherwise, the charge-generation efficiency at selected field is given. Although the Onsager model has its limitations and problems, it nevertheless provides a common platform to compare various photoconductors, since many important photoconductors have been analyzed by this model. Not much significance should be attached to the molecular meaning of r_0 .

The intrinsic charge-generation efficiency of polymers is often low and needs to be enhanced by the addition of sensitizers. The sensitizer can be dissolved in the polymer to enhance the bulk charge-generation efficiency of the polymer. Effective sensitizers include 2,4,7-trinitro-9-fluorenone [129-79-3] (TNF), fullerene, thiapyrylium dye, CdS nanoclusters, etc (Table 3). Molecular structures of selected sensitizers are shown in Figure 8.

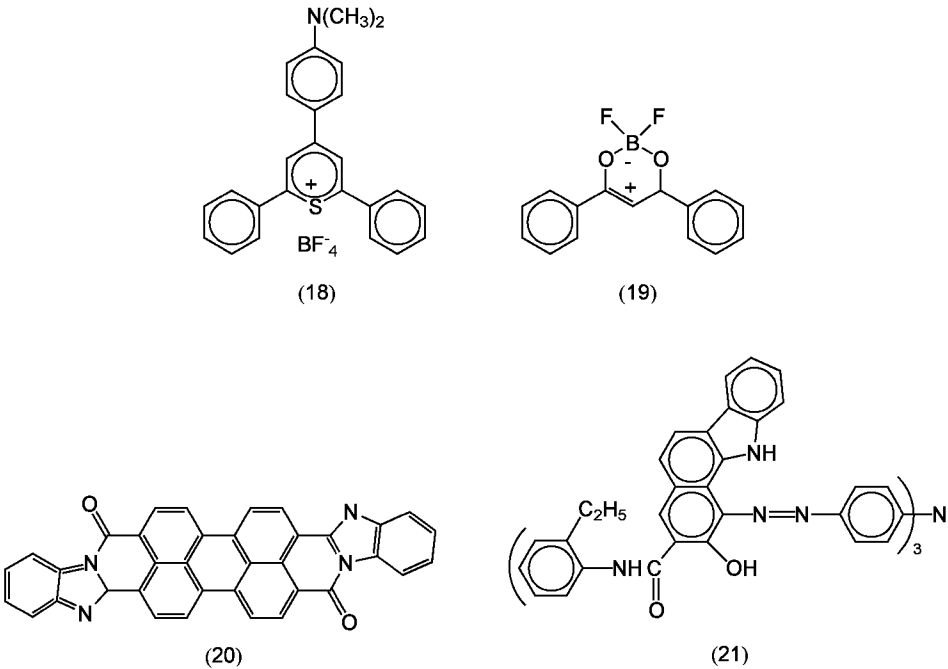


Fig. 8. Molecular structures of selected sensitizers: (18) thiapyrylium dye [25966-12-5]; (19) difluoroboron-1,3-diphenyl-1,3-propanedionate (DPBDK); (20) benzimidazole (BZ) perylene pigment; and (21) triphenylamine trisazo pigment (TPATA).

The charge-generation efficiency of a polymer can also be sensitized externally in a bilayer device. In that case, a sensitizer dye is deposited as a charge-generation layer on top of a polymer film acting as a charge-transport layer. The overall charge-generation efficiency depends not only on the intrinsic charge-generation efficiency of the sensitizer, but also on the electron-transfer efficiency across the interface from the sensitizer to the polymer. It was found that the total charge-generation efficiency of such a bilayer device is often higher than that of the charge-generation layer alone (eg, Table 3) (79,80). Charge-generation efficiency as high as ca 0.8 at 8×10^5 V cm has been achieved in a TPATA bilayer device (Table 3). By using TPATA as the charge-generation layer and a series of amines with different redox potential as the charge-transporting layer, it was found that the charge-generation efficiency depends on the free-energy difference of the interfacial electron-transfer reaction between TPATA and the amines (84). The dependence can be described by the Marcus electron-transfer theory. This result clearly confirms the importance of interfacial electron-transfer in a bilayer device.

Applications and Related Technologies

The most important industrial application of photoconductive polymers is electrophotography (qv). This is a billion dollar industry and one of the few electronic areas where polymeric material excels. The principles and practices of electrophotography have been reviewed in detail elsewhere (9,85) and are not repeated here.

The availability of photoconductive polymers opens up many areas for research, in addition to electrophotography. These are relatively unexplored areas and represent promising future directions.

Electroluminescence. Photoconductivity is based on the conversion of light to electricity. The reverse phenomenon, electroluminescence, is based on the conversion of electricity to light. Electroluminescence is useful for flat-panel display and II–VI semiconductors such as ZnS are employed for this purpose (85). The current trend is toward the development of polymeric electroluminescent material for their processing flexibility (86–88). Hole-transporting polymers such as poly(*p*-phenylenevinylene) (87) and PMPS (88) have been used in such devices. Semiconductor nanocluster-doped polymers represent another interesting class of materials for the exploration of electroluminescent phenomenon. It has already been demonstrated that properly doped semiconductor nanoclusters such as Zn_xMn_{1-x}S emits light efficiently (89). With the demonstration of photoconductivity (26) these nanocluster-doped polymers can become possible candidates for electroluminescent materials (90).

Photorefractive Effect. If a material possesses second-order optical nonlinearity and is photoconductive, it may be photorefractive (91). Photorefractivity is a third-order nonlinear optical, χ^3 , phenomenon. The phenomenon occurs as photogenerated carriers redistribute and create an inhomogeneous space-charge field in the medium. If the material has second-order optical nonlinearity, χ^2 , this space-charge field can modulate the refractive index of the material through the electrooptical effect. Photorefractive materials provide a medium in which holographic gratings can be reversibly written and are useful as optical interconnects. A good photorefractive material requires large second-order nonlinearity and high charge-generation efficiency.

Recently photorefractivity in photoconductive polymers has been demonstrated (92–94). The second-order nonlinearity is obtained by poling the polymer doped with a nonlinear chromophore. Such a polymer may or may not be a good photoconductor. Usually sensitizers have to be added to enhance the charge-generation efficiency. The sensitizer function of fullerene in a photorefractive polymer has been demonstrated (93).

Data Storage. An interesting extension of photoconductor technology to optical data storage has been reported (95). The device consists of a solid thin film of the photoconductor, zinc-octakis(β-decoxyethyl)porphyrin, sandwiched between two transparent electrodes. Irradiation by light (550 nm) under an applied electric field generates electron–hole pairs which are separated within the photoconductive layer. When the irradiation is interrupted, these electron–hole pairs become trapped within the film because of the low dark conductivity. This corresponds to the data-writing step. The written information can be read by irradiation of the device with a read beam (at 550 nm) under short circuit conditions. The release of electrons or holes from the traps leads to a photocurrent spike, which can be taken to represent the memory state, "1," compared with an uncharged state, "0." The basic principle behind this data storage scheme is similar to that of xerography, where the readout method is by toning with carbon particles to form an image (96). The reported method represents a digital way of reading the electrostatic image (95) (see INFORMATION STORAGE MATERIALS, OPTICAL).

This new optical data storage device is reported to be robust and nonvolatile. The response time for the write–read beam is in the subnanosecond range, and no refreshing is required for long-term retention of trapped charges (95). The basic principle may be applied to other, similar photoconductive materials.

BIBLIOGRAPHY

1. R. B. Seymour, ed., *Conductive Polymers*, Plenum Press, New York, 1981.

2. **U.S. Pat. 3,037,861** (June 5, 1962), H. Hoegl, O. Sus, and W. Neugebauer (to Hoechst AG).

3. **U.S. Pat. 3,484,237** (Dec. 16, 1969), D. M. Shattuck and U. Vahtra (to IBM).

4. D. Haarer, *Angew. Makromol. Chem.* **183**, 197 (1990).

5. M. Stolka, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, Vol. 11, Wiley-Interscience, New York, 1987, p. 154.

6. W. D. Gill, in J. Mort and D. M. Pai, eds., *Photoconductivity and Related Phenomena*, Elsevier, Amsterdam, the Netherlands, 1976, p. 303.

7. H. Bassler, *Phys. Stat. Sol.(b)* **175**, 15 (1993).

8. J. Mort and G. Pfister, in J. Mort and G. Pfister, eds., *Electronic Properties of Polymers*, John Wiley & Sons, Inc., New York, 1982, p. 215.

9. L. B. Schein, *Electrophotography and Development Physics*, Springer-Verlag, Bedin, 1988.

10. M. Biswas and T. Uryu, *J. Macromol. Sci., Rev. Macromol. Chem. Phys.* **C26**, 248 (1986).

11. W. D. Gill, *J. Appl. Phys.* **43**, 5033 (1972).

12. Y. Wang, *Nature* **356**, 585 (1992).

13. Y. Wang, R. West, and C. H. Yuan, *J. Amer. Chem. Soc.* **115**, 3844 (1993).

14. R. West, in S. Patai and Z. Rappoport, eds., *The Chemistry of Organic Silicon Compounds*, John Wiley & Sons, Inc., New York, 1989, Chapt. 19, pp. 1207–1240.

15. R. D. Miller and J. Michl, *Chem. Rev.* **89**, 1359 (1989).

16. R. G. Kepler, J. M. Zeigler, L. A. Harrah, and S. R. Kurtz, *Phys. Rev. B*, **35**, 2818 (1987).

17. M. Abkowitz, F. E. Knier, H.-J. Yuh, R. J. Weagley, and M. Stolka, *Solid St. Comm.* **62**, 547 (1987).

18. M. Stolka, H.-J. Yuh, K. McGrane, and D. M. Pai, *J. Polym. Sci. Polym. Chem. Ed.* **25**, 823 (1987).

19. M. Fujino, *Chem. Phys. Lett.* **136**, 451 (1987).

20. M. A. Abkowitz, K. M. McGrane, F. E. Knier, and M. Stolka, *Mol. Cryst. Liq. Cryst.* **183**, 157 (1990).

21. S. Morita, S. Kiyomatsu, X. H. Yin, A. A. Zakhidov, T. Noguchi, T. Ohnishi, and K. Yoshino, *J. Appl. Phys.* **74**, 2860 (1993).

22. L. L. Chapoy and co-workers, *Mol. Cryst. Liquid Cryst.* **105**, 353 (1984).

23. L. L. Chapoy and co-workers, *Macromolecules* **16**, 181 (1983).

24. D. Adam and co-workers, *Phys. Rev. Lett.* **70**, 457 (1993).

25. J. Billard, J. C. Dubois, N. H. Tinh, and A. Zahn, *Nouv. J. Chim.* **2**, 535 (1978).

26. Y. Wang and N. Herron, *Chem. Phys. Lett.*, **200**, 71 (1992).

27. **U.S. Pat. 5,238,607** (Aug. 24, 1993), N. Herron, Y. Wang (to Du Pont Co.).

28. Y. Wang and N. Herron, *J. Phys. Chem.* **95**, 525 (1991).

29. Y. Wang, in D. C. Neckers, ed., *Advances in Photochemistry*, Vol. 19, John Wiley & Sons, Inc., New York, 1995.

30. Y. Wang and N. Herron, *Res. Chem. Intermedi.* **15A**, 17 (1991).

31. O. H. Le Blanc, *J. Chem. Phys.* **33**, 626 (1960).

32. K. G. Kepler, *Phys. Rev.*, **199** 1226 (1960).

33. F. K. Dolezalek in J. Mort and D. M. Pai, eds., *Photoconductivity and Related Phenomena*, Elsevier, Amsterdam, the Netherlands, 1976, p. 27.

34. H. Scher and M. Lax, *Phys. Rev. B* **7**, 4491, 4502 (1973).

35. H. Scher and E. W. Montroll, *Phys. Rev. B* **12**, 2455 (1975).

36. H. Bassler, *Philos. Mag. B* **50**, 347 (1984).

37. P. M. Borsenberger, L. Pautmeier, H. Bassler, *J. Chem. Phys.* **94**, 5447 (1991).

38. B. Movaghar, M. Gr̃newald, B. Ries, H. Bassler, and D. W̃rtz, *Phys. Rev. B* **33**, 5545 (1986).

39. P. M. Borsenberger, L. Pautmeier, and H. Bassler, *Phys. Rev. B* **46**, 12145 (1992).

40. D. Emin, in P. G. LeComber and J. Mort, eds., *Electronic and Structural Properties of Amorphous Semiconductors*, Academic Press, Inc., New York, 1973, Chapt. 7.

41. L. B. Schein, D. Glatz, and J. C. Scott, *Phys. Rev. Lett.* **65**, 472 (1990).

42. L. B. Schein, *Philos. Mag. B* **65**, 795 (1992).

43. P. M. Borsenberger and L. B. Schein, *J. Phys. Chem.* **98**, 233 (1994).

44. D. M. Pai, *J. Chem. Phys.* **52**, 2285 (1970).

45. Technical data, Du Pont Co., 1995.

46. T. Sasakawa, T. Ikeda, and S. Tazuke, *Macromol.* **22**, 4253 (1989).

47. H. Domes, R. Fischer, D. Haarer, and P. Strohriegl, *Makromol. Chem.* **190**, 165 (1989).

48. M. Stolka and M. A. Abkowitz, *Syn. Met.* **54**, 417 (1993).

49. D. M. Pai, J. F. Yanus, and M. Stolka, *J. Phys. Chem.* **88**, 4714 (1984).

50. J. S. Facci and co-workers, *J. Phys. Chem.* **95**, 7908 (1991).

51. A. Yu. Kryukov, A. A. Pakhratdinov, A. V. Vannikov, A. V. Anikeev, L. I. Kostendo, *Chem. Mater.* **4**, 72 (1992).

52. K. Yokoyama, S. Notsu, and M. Yokoyama, *J. Chem. Soc., Chem. Commun.*, 805 (1990).

53. H. Bassler, *Adv. Mat.* **5**, 662 (1993).

54. H.-J. Yuh and D. M. Pai, *Mol. Cryst. Liquid Cryst.* **183**, 217 (1990).

55. L. B. Schein, A. Rosenberg, and S. L. Rice, *J. Appl. Phys.* **60**, 4287 (1986).

56. P. M. Borsenberger, W. Mey, and A. Chowdry, *J. Appl. Phys.* **49**, 273 (1978).

57. G. Pfister, *Phys. Rev. B* **16**, 3676 (1977).

58. D. M. Pai, J. F. Yanus, M. Stolka, D. Renfer, and W. W. Limburg, *Phil. Mag. B*, **48**, 505 (1983).

59. P. M. Borsenberger, A. Chowdry, D. C. Hoesterey, and W. Mey, *J. Appl. Phys.* **49**, 5555 (1978).
60. T. Ikeda, H. Mochizuki, Y. Hayashi, and M. Sisido, *J. Appl. Phys.* **70**, 3689 (1991).
61. P. M. Borsenberger and J. J. Fitzgerald, *J. Phys. Chem.* **97**, 4815 (1993).
62. P. M. Borsenberger and H. Bassler, *J. Chem. Phys.* **95**, 5327 (1991).
63. P. M. Borsenberger, J. J. Fitzgerald, and E. H. Magin, *J. Phys. Chem.* **97**, 11314 (1993).
64. P. M. Borsenberger, E. H. Magin, and J. J. Fitzgerald, *J. Phys. Chem.* **97**, 8250 (1993).
65. M. Stolka, J. F. Yanus, and D. M. Pai, *J. Phys. Chem.* **88**, 4707 (1984).
66. M. Van der Auweraer, F. C. DeSchryver, P. M. Borsenberger, and J. J. Fitzgerald, *J. Phys. Chem.* **97**, 8808 (1993).
67. P. J. Regensburger, *Photochem. Photobiol.* **8**, 429 (1968).
68. I. Chen and J. Mort, *J. Appl. Phys.* **43**, 1164 (1972).
69. Z. D. Popovic and E. R. Menzel, *J. Chem. Phys.* **71**, 5090 (1979).
70. Z. D. Popovic, *J. Chem. Phys.* **78**, 1552 (1983).
71. M. Yokoyama, S. Shimokihara, A. Matsubara, and H. Mikawa, *J. Chem. Phys.* **76**, 724 (1982).
72. L. Onsager, *Phys. Rev.* **54**, 554 (1938).
73. A. Mozumder, *J. Chem. Phys.* **60**, 4300 (1974).
74. G. Pfister and D. J. Williams, *J. Chem. Phys.* **61**, 2416 (1974).
75. P. J. Melz, *J. Chem. Phys.* **57**, 1694 (1972).
76. C. F. Hackett, *J. Chem. Phys.* **55**, 3178 (1971).
77. P. M. Borsenberger and D. C. Hoesterey, *J. Appl. Phys.* **51**, 4248 (1980).
78. T. E. Goliber, J. H. Perlstein, *J. Chem. Phys.* **80**, 4162 (1984).
79. Z. D. Popovic, A.-M. Hor, and R. O. Loutfy, *Chem. Phys.* **127**, 451 (1988).
80. M. Umeda and M. Hashimoto, *J. Appl. Phys.* **72**, 117 (1992).
81. C. L. Braun, *J. Chem. Phys.* **80**, 4157 (1984).
82. J. Noolandi and K. M. Hong, *J. Chem. Phys.* **70**, 3230 (1979).
83. R. A. Marcus and P. Siders, *J. Phys. Chem.* **86**, 622 (1982).
84. M. Umeda, T. Shimada, T. Aruga, T. Niimi, and M. Sasaki, *J. Phys. Chem.* **97**, 8531 (1993).
85. R. Mach and G. O. Muller, *Phys. Stat. Sol.* **69**, 11 (1982).
86. C. W. Tang and S. A. Vanslyke, *Appl. Phys. Lett.* **51**, 913 (1987).
87. J. H. Burroughes and co-workers, *Nature* **347**, 539 (1990).
88. H. Suzuki, H. Meyer, J. Simmerer, J. Yang, and D. Haarer, *Adv. Mater.* **5**, 743 (1993).
89. Y. Wang, N. Herron, K. Moller, and T. Bein, *Solid State Comm.* **77**, 33 (1991).
90. B. O. Dabbousi and co-workers, *Appl. Phys. Lett.* **66**, 1316 (1995).
91. P. Gunter and J.-P. Huignard, eds., *Photorefractive Materials and Their Applications I and II*, Springer-Verlag, Berlin, 1988, 1989.
92. S. Ducharme, J. C. Scott, R. J. Twieg, and W. E. Moerner, *Phys. Rev. Lett.* **66**, 1846 (1991).
93. S. M. Silence, C. A. Walsh, J. C. Scott, and W. E. Moerner, *Appl. Phys. Lett.* **61**, 2967 (1992).
94. J. C. Scott, L. T. Pautmeier, and W. E. Moerner, *Synth. Met.* **54**, 9, 1993.
95. C.-Y. Liu, H.-L. Pan, M. A. Fox, and A. J. Bard, *Science*, **261**, 897 (1993).
96. R. M. Schaffert, *Electrophotography*, Focal Press, London, 1980.

Ying Wang
Du Pont Company

PHOTODEGRADABLE POLYMERS.

See POLYMERS, ENVIRONMENTALLY DEGRADABLE

PHOTODETECTORS

Photons impinging upon matter interact in a manner determined by the nature of the chemical bonds in the material and the energy of the incident photons. Interacting photons may be reflected, refracted, diffracted, transmitted, or absorbed. Each of these phenomena can be used to measure some parameter of interest in chemical analysis. Photo-assisted chemical analytical techniques require an accurate, sensitive method of photon detection and quantification (see ANALYTICAL METHODS). Photographic film proved to be the first useful method for such evaluations (see PHOTOGRAPHY). Later, photomultiplier tubes provided the means to improve many photo-analytical techniques. With the advent of semiconductor technology, numerous single-element, broad-spectrum photodetecting devices were developed and applied to the task (see SEMICONDUCTORS). Most recently photodetector technology has entered the age of high density integration. Large-area photon detectors are becoming commonplace in most analytical equipment. The advances afforded by these devices have allowed for significant improvements in the performance of systems designed for photo-analytical chemical analysis in general, and specifically for spectroscopic-based equipment. A working knowledge of the operation and limitations of photodetectors is therefore necessary for the modern chemist.

Photodetector devices convert electromagnetic radiation or photons to electric signals which can be processed to obtain the spectral, spatial, and temporal information inherent in the radiation. Photodetectors, as indicated in Table 1, may be operated in many modes. The more popular ones are photoconductors, photodiodes, charge-transfer devices, and the pyroelectrics. The detectors may be used as single elements such as in street light controls, film camera exposure control, or motion detectors for security, or in the form of linear arrays used in analytical spectrometers, night-vision equipment, or configured as large matrix arrays found in video cameras. By the twenty-first century, photodetectors are expected to appear in small, low cost spectrometers for the control of building ventilation and environmental pollution monitoring. Detectors using artificially structured materials such as semiconductor superlattices and high temperature superconductors are in the early development stage as of this writing (ca 1995).

Table 1. Semiconductor Photodetectors in General Use

Bandwidth , μm	Mode a,b	Temp., K	Material	Pixel size, μm	Focal plane size	Principle use	Cutoff wavelengt h, μm	Response time, μs	Responsivi ty, V W	D^* , cm-Hz ^{1/2} W
0.200–0.400	CCD	300	Si	10–30	400 × 400	astronomy	0.9		2 × 10 ¹⁰	1 × 10 ¹³
0.400–0.700 ^c	CCD	300	Si	10–30	488 × 640–2000 × 2000	video camera	1.0		1 × 10 ¹¹	5 × 10 ¹³
	PC	300	CdS	100–1000	<100	light meter	0.7	100,000	1 × 10 ⁶	1 × 10 ¹³
	PV	300	Si	30–200	few	optical switch	1.0	0.01–10	0.5 ^d	5 × 10 ¹³
	PV	300	GaAsP	30–200	few	optical switch	0.8	0.1–10	0.2 ^d	1 × 10 ¹³
1–3	PV	240	InGaAs	30–200	1 × 300	spectroscopy	1.7	1	0.2 ^d	2 × 10 ¹²
	PV	200–300	HgCdTe	40–400	10 × 300	mapping	2.9	10	2.0 ^d	1 × 10 ¹²
	PC	200–300	PbS	40–400	10 × 3000	mapping	3.3–2.7	1000–300	5 × 10 ⁴ – 1 × 10 ³	5 × 10 ¹¹ – 1 × 10 ¹¹
3–5	PC	150–220	HgCdTe	40–60	1 × 100	night scope; heat sensing	5.0	10	1 × 10 ⁶	1 × 10 ¹¹
	PC	77–200	PbSe	50–80	1 × 100	heat sensing	5.8 ^e	150 ^e	1 × 10 ⁶ ^e	2 × 10 ¹⁰ ^e
	PC ^f	60–90	Ge: Au	80–200	few	fast response	7.8	0.1	1 × 10 ⁴	1 × 10 ¹⁰
	PV	60–90	InSb	40–60	256 × 256	thermal imaging	5.25	1	3.4 ^d	1 × 10 ¹¹
	PV	120	HdCdTe	40–60	256 × 256	thermal imaging	5.0	5	3.4 ^d	1 × 10 ¹¹
	SB ^g	60–80	Si:Pt	40–60	480 × 640	thermal imaging	4.6	1	0.03 ^d	3 × 10 ⁹
8–12	PC	70–100	HgCdTe	40–60	1 × 200	thermal imaging	12.2	1	1 × 10 ⁶	3 × 10 ¹⁰
	PV	60–90	HgCdTe	40–60	256 × 256	thermal imaging	10.5	1	6.5 ^d	5 × 10 ¹⁰
	QWI	50–77	GaAs A	40–60	128 × 128	thermal	9.4 ⁱ	1 ⁱ	0.2 ^{d,i}	1 × 10 ¹⁰ ⁱ

5–150	p ^h		1GaAs							
	PC ^f	30–40	Ge:Hg	50–200	1 × 200	imaging thermal	13	0.1	5 × 10 ⁵	3 × 10 ¹⁰
						imaging spectrome	28 ^l	1 ^l	5 × 10 ^{5.1}	3 × 10 ^{10.1}
	PC ^f	20–30	Ge:Cd	50–200	1 × 100	try spectrome	37 ^k	1 ^k	5 × 10 ^{5.k}	2 × 10 ^{10.k}
	PC ^f	5–15	Ge:Zn	50–200	1 × 100	try threat	18	1	5 × 10 ⁵	2 × 10 ¹⁰
0.200–200	PC ^f	4–10	Si:Ga	50–100	64 × 256	detection threat	22	1	5 × 10 ⁵	2 × 10 ¹⁰
	PC ^f	4–10	Si:As	50–100	64 × 256	detection				
0.200–200	Bol ^l	300	α-silicon	40–60	240 × 320	thermal	full band	10,000	1 × 10 ⁶	5 × 10 ⁸
	Bol ^l	300	VO	40–60	240 × 320	imaging thermal	full band	10,000	1 × 10 ⁶	5 × 10 ⁸
						imaging				

^a Bol = bolometer; CCD = charge-coupled device; PC^f = photoconductor; PV = photovoltaic; QWIP = quantum well infrared photodetector; and SB = Schottky barroer.

^b Detectors are all intrinsic unless otherwise noted. See Fig. 1.

^c Visible bandwidth; near infrared = 0.700–1.00 μm.

^d Units are A W.

^e Values are for a temperature of 190 K.

^f Detector is extrinsic.

^g Detector is an internal photo-emission (IPE) device.

^h Detector is a superlattice (SL) device.

ⁱ Values are for a temperature of 77 K.

^j Values are for a temperature of 15 K.

^k Values are for a temperature of 6 K.

^l Detector is a thermal device.

Principles

Photon-Absorption and Thermal Emission. The basic detection process, as shown in Figure 1, is the generation of free electrons, holes, or both by the absorption of photon energy (1–4). The absorption may be intrinsic, ie, creating a free electron–hole pair, or extrinsic, creating a free hole or electron. The process can be indirect whereby the absorbed photon energy raises the lattice temperature (thermal detection) and a phonon generates the free charged particle. The change in free charge is sensed in an amplifier circuit as a signal voltage or current. Random generation of free charge is the source of detector noise. The detector geometry, spectral response, electrical bias, and temperature are adjusted to optimize the signal-to-noise ratio. The absorption efficiency (photon to electron conversion) is a critical issue. The detector surface is typically coated to minimize reflections at a particular wavelength and the internal efficiency is given by $1 - \exp(-\alpha x)$. Because the absorption coefficient α is typically 3000 cm⁻¹ for intrinsic detection and 3 cm⁻¹ for the extrinsic case, detector thickness, x , has a large range of values.

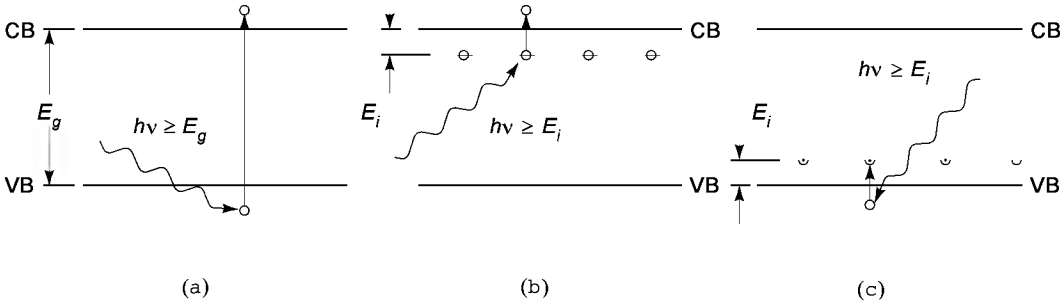


Fig. 1. Photoexcitation modes in a semiconductor having band gap energy, E_g , and impurity states, E_i . The photon energy $h\nu$ must be sufficient to release an electron ($\circ$) into the conduction band (CB) or a hole (ω) into the valence band (VB): (a) an intrinsic detector; (b) and (c) extrinsic donor and acceptor devices, respectively.

There are important figures of merit (5) that describe the performance of a photodetector. These are responsivity, noise, noise equivalent power, detectivity, and response time (2,6). However, there are several related parameters of measurement, eg, temperature of operation, bias power, spectral response, background photon flux, noise spectra, impedance, and linearity. Operational concerns include detector-element size, uniformity of response, array density, reliability, cooling time, radiation tolerance, vibration and shock resistance, shelf life, availability of arrays, and cost.

Photodetectors exhibit well-defined, cutoff wavelength thresholds, the positions of which are determined by the magnitudes of the band gap activation energy, E_g , or impurity-activation energy, E_i . The cutoff wavelength, λ_c , corresponds to a photochemical activation energy, E , where,

$$\lambda_c = \frac{hc}{E}$$

(1)

The product of Planck's constant and the velocity of light, $hc = 1.24$ when E is expressed in electron volts and λ_c in μm. Because the absorption coefficient typically is a weak function of wavelength above the threshold energy, the spectral response approximates that of the ideal quantum counter. The interest in semiconductor photodetectors with cutoff wavelengths longer than 3 μm results from their extreme sensitivity to variations in photon flux emitted from gray-body objects that are characterized by time and spatial variations in temperature and emissivity. Although some applications require detection of narrow-band radiation, eg, atomic emission spectra and radiation from hot gases, most uses of infrared detectors involve detection of variations in photon flux from objects near ambient temperature. Typically, a lens system images the scene of interest on a single detector element, a linear array of detector

elements, or an area array (7,8). The infrared detector is exposed to the signal photon flux through the optics. The ambient background photon flux, ϕ_{B} is often much greater than the signal flux and determines the ultimate system performance. The background portion, ϕ_{B} in photons (cm²s) may be calculated from Planck's radiation law.

Blackbody Emittance. Representative blackbody emittance (9,10), calculated as a power spectral density, is shown in Figure 2. The wavelength, λ , of peak power density for a blackbody at temperature, T , is given by Wien's displacement law:

$$\lambda_m T = 2898 \text{ }\mu\text{mK}$$

(2)

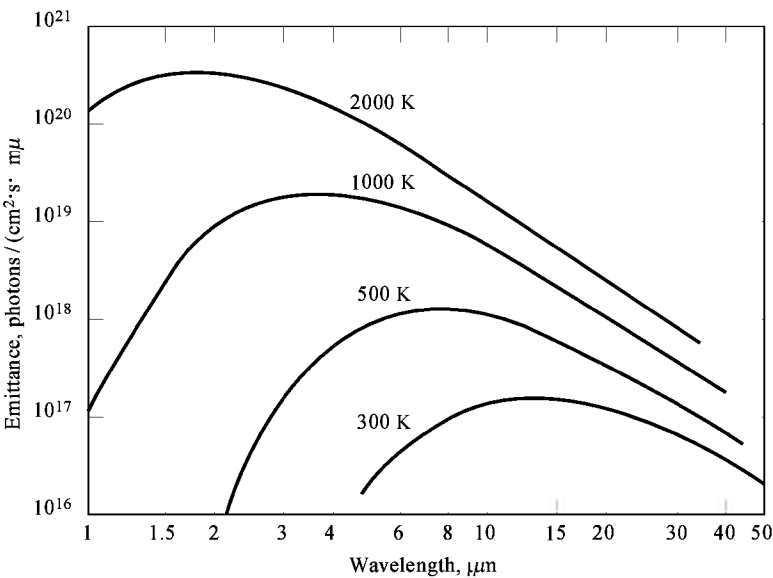


Fig. 2. Blackbody radiated photon flux interval distribution from ambient temperature up to 2000 K. Ambient objects do not have detectable flux for detectors of cutoff wavelengths <ca 3 μm .

where λ_m is the wavelength of maximum power density and T is temperature in K. The peak spectral power density or emittance, in W (cm²· μm), is given by

$$M(T)_m = 1.288 \times 10^{-15} T^5$$

(3)

for each square centimeter of emitting surface and μm wavelength at the peak value. The total power density, W , as integrated over all wavelengths in W cm², is given by the Stefan Boltzmann law:

$$W = 5.67 \times 10^{-12} T^4$$

(4)

which expresses the total rate of emission in a 2π steradian solid angle from an element area on a blackbody surface. The very strong dependence of W on temperature and the sharp decline of emittance for wavelengths somewhat less than the peak wavelength results in very low ambient background flux density for ambient temperatures less than 320 K and detectors having cutoff wavelengths <3 μm . Short wavelength detectors such as silicon charge-coupled devices (CCDs) do not work in the infrared for this reason but are excellent when hot sources are available such as light bulbs (1500 K) and the sun (6000 K). It is apparent from Figure 2 that detection of ambient temperature objects (night vision) requires a detector having cutoff wavelength >4 μm . Detectors having response beyond 8 μm generally are better for thermal imaging applications. However, system sensitivity is dependent also on several other factors: detector quantum efficiency, detector noise, electronics noise, number of detectors on the focal plane (ie, noise bandwidth), sampling rate, width of spectral band, optical aperture, and detector element size.

Figures of Merit

Responsivity. A detector responds to variations in photon-flux density ϕ_i by producing a signal voltage, V_s , or current. The background photon flux, ϕ_{B} , may result in an offset voltage which can be nulled by a differentiating circuit, because ϕ_{B} is constant for a given measurement. For most detectors, the output voltage depends on the additive effect of ϕ_i over the detector area A . Each photon has energy $h\nu$. The integrated power density over a spectral interval is J_i . The ratio of signal voltage (amperes or electrons) to the signal photon rate is the responsivity, R_{V_i} , in units of V W.

$$R_V = \frac{V_s}{A J_s}$$

(5)

V_i usually is linear with J_i for low intensity, but saturates at the higher signal-flux levels where the photon flux disrupts the thermal equilibrium of the semiconductor electronic states. When J_i refers only to those photons capable of generating free carriers, R_{V_i} is considered the spectral responsivity, R_{λ_i} ; when the flux includes all wavelengths, R_{V_i} is called the blackbody responsivity, R_{bb} . The responsivity of photodiodes is often expressed in amperes per watt (A W) because the diode is utilized as a current generator when connected to an integrating amplifier. The ideal responsivity for a photodiode is given by $\eta q/E_g$ where η is the quantum efficiency, q is the charge of an electron, and E_g is the band gap energy. Most commercial detectors achieve greater than 50% of the ideal responsivity except for the silicide detectors which typically show less than 2% of the ideal responsivity.

Noise. Performance of a system, whether it be an analytical spectrometer or thermal imager, improves as the detector noise is lowered until amplifier noise becomes the dominant noise source, provided that the responsivity is not reduced faster than the detector noise. A well-engineered approach to improvement of system sensitivity consists of (1) lowering amplifier noise consistent with system bandwidth, power, weight, size, and cost budgets; (2) perfecting the detector material and fabrication processes for least defects to achieve lowest noise; and (3) designing the detector mode of operation, geometry, response time, and bias power for highest responsivity. The operating parameters of the detector are often adjusted to increase responsivity and noise to the point where the detector noise just exceeds the amplifier noise if in fact that is possible.

Although photodetector noise (11) originates as random fluctuations of photons or electrical carriers, it is measured using the Fourier transform

process into a frequency function for purposes of analysis. The noise spectrum is characterized by frequency dependent noise at low frequency (called 1/f noise), generation recombination (g-r) noise in a midrange, and system noise at high frequency. At low frequency, noise typically varies inversely as the square root of frequency and is a result of surface traps and/or fluctuating defects (12,13). This noise is often reduced by surface passivation to reduce the trap density or create a surface potential to keep the charge carriers away from the surface. Crystal growth resulting in fewer electronic defect states (typically $<1 \times 10^{15} \text{ cm}^{-3}$) and elimination of detector fabrication damage also reduces 1/f noise. A well-designed detector is dominated by g-r noise associated with random carrier generation and recombination in the semiconductor. The minority carrier lifetime, τ , determines the extent of the g-r noise plateau. Trapping mechanisms, whether bulk or surface, exhibit behavior similar to the g-r process when the trapping time determines the detector response time. Careful analysis is required to distinguish between recombination and trapping. This identification helps in determining the nature of the defect such as the identity of the impurities. For example, copper atoms in HgCdTe reduce the lifetime and detector responsivity. Characterized by τ , the g-r or trapping noise declines as the inverse of frequency and the system or Johnson noise (JN) becomes dominant at the higher frequencies. In general, noise components, V_p , are not correlated adding in quadrature to give the total noise as follows:

$$V_N^2 = V_f^2 + V_{g-r}^2 + V_{JN}^2 + V_{AMP}^2$$

(6)

The noise is expressed as noise density in units of $V \text{ (Hz)}^{-1/2}$, or integrated over a frequency range and given as volts rms. Typically, photoconductors are characterized by a g-r noise plateau from 10^3 to 10^5 Hz. Photovoltaic detectors exhibit similar behavior, but the 1/f knee may be less than 100 Hz and the high frequency noise roll off is determined by the *p-n* junction impedance–capacitance product or the amplifier (AMP) circuit when operated in a transimpedance mode. Bolometers exhibit an additional noise, V_{TC} , associated with thermal conductance.

Detectivity. Detector sensitivity (1,2) is expressed in terms of the minimum detectable signal power or noise equivalent power (NEP) given in units of watts or $W \text{ Hz}^{-1/2}$. The reciprocal function when normalized for detector area, *A*, and noise bandwidth, Δf is defined as detectivity, D^* , in units of $\text{cm}^2\text{Hz}^{1/2} \text{ W}^{-1}$. Thus,

$$\text{NEP} = \frac{V_N}{R_V}$$

(7)

$$D^* = \frac{R_V (A \Delta f)^{1/2}}{V_N}$$

(8)

The rms noise is measured in a noise bandwidth, Δf . The D^* is called D star lambda when the spectral band is limited to a given interval, and D^* blackbody when the total blackbody incident power density is used in the calculation.

The blackbody method of calibrating photodetectors is excellent for cutoff wavelengths greater than 2 μm , but for short wavelength detectors such as visible sensing charge-coupled devices the unit of lux is the scattered illumination power density of $100 \mu\text{W cm}^{-2}$ in the visible spectrum. Most CCD cameras have a sensitivity of 3–10 lx. Spatial resolution and image display is often a problem with arrays of discrete detectors because the small objects are undersampled. The results are Moiré patterns and loss of detail. Scanning arrays use overlapping detector geometries to improve the spatial quality of the image. Infrared detector element size is limited to greater than 10 μm by electrical contact methods or photolithographic minimum feature size (1 μm) for the detectors or readout electronics.

Ideal Performance and Cooling Requirements. Free carriers can be excited by the thermal motion of the crystal lattice (phonons) as well as by photon absorption. These thermally excited carriers determine the magnitude of the dark current, J_d , and constitute a source of noise that defines the limit of the minimum radiation flux that can be detected. The dark carrier concentration is temperature dependent and decreases exponentially with reciprocal temperature at a rate that is determined by the magnitude of E_g or E_i for intrinsic or extrinsic material, respectively. Therefore, usually it is necessary to operate infrared photon detectors at reduced temperatures to achieve high sensitivity. The smaller the value of E_g or E_i , the lower the temperature must be.

The operating temperature depends on the required sensitivity and the temperature dependence of the various thermal–noise components. In all cases, a practical limitation in sensitivity is imposed by an amplifier or recording device. Theoretical performance can be calculated, assuming that the amplifier noise is negligible as compared with detector noise. Optimum or ideal detector performance is based on the assumption that the detector noise is determined only by the random fluctuations of the photon-absorption process. This ultimate limit on performance can be calculated easily and is used as a final measure of performance for infrared detectors (14). For example, consider a photodetector as a photon counter, such as a photodiode connected to an integrating amplifier. After a time interval, Δt , the integrated charge density, Q , may be expressed by the following:

$$Q = (J_\phi + J_d) \Delta t$$

(9)

The number of integrated carriers, N , is QA/q , where q is the electron charge. Because dark current, J_d , is a combination of thermal excitation processes, neglecting avalanche and tunneling, ideal performance occurs when the photon-induced current density J_ϕ is greater than J_d . Fluctuations of N are the dominant causes of noise. The cooling requirement may be calculated for each type of photon detector by analyzing the dark current noise mechanism in terms of thermal equilibrium processes and setting $J_\phi > J_d$ with $J_\phi = \eta \phi q A \text{ cm}^{-2}$ and ϕ is the sum of the signal and background photon flux. Because the charge density of a photon-counting detector is sampled at time intervals, Δt , the number of charges sampled $N = \eta \phi A \Delta t$.

The generation of photons obeys Poisson statistics where the variance is N and the deviation or noise is $N^{1/2}$. The noise spectral density, N_s , is obtained by a Fourier transform of the deviation yielding the following at sampling frequency, f :

$$N_s = \left(\frac{2N\Delta f}{f} \right)^{1/2} = \left(\frac{2\eta \phi A \Delta t \Delta f}{f} \right)^{1/2}$$

(10)

The signal S , in electrons, is that portion of N which is generated by a signal flux, ϕ_s , in the same time interval, Δt . Therefore,

$$S = \eta \phi_s A \Delta t$$

(11)

The ideal detectivity may be obtained by combining equations 5, 8, 10, and 11. When $\Delta t = f^{-1}$, the ideal spectral detectivity (2) in $\text{cm}^2\text{Hz}^{1/2} \text{ W}^{-1}$ for a photon limited process becomes:

$$D_\lambda^* = \frac{\lambda}{hc} \left(\frac{\eta}{2\phi} \right)^{1/2}$$

(12)

When equation 12 is valid, the detector is said to be a background-limited infrared photodetector (BLIP). When this is the case, attempts often are made to improve D^* by cold shielding which reduces ϕ . The ideal D^* is shown in Figure 3 as a function of wavelength with background photon flux as a parameter. The line of termination in the lower left corner represents D^* values for a 180° (2π) detector field of view, 300 K ambient background

temperature, and background emissivity $\epsilon = 1.0$. As long as the photon-generated current exceeds the dark currents, D^* may be increased by reducing the background flux. For very low values of photon flux, the ideal detection condition of equation 12 can be restored by cooling the detector to decrease the dark currents.

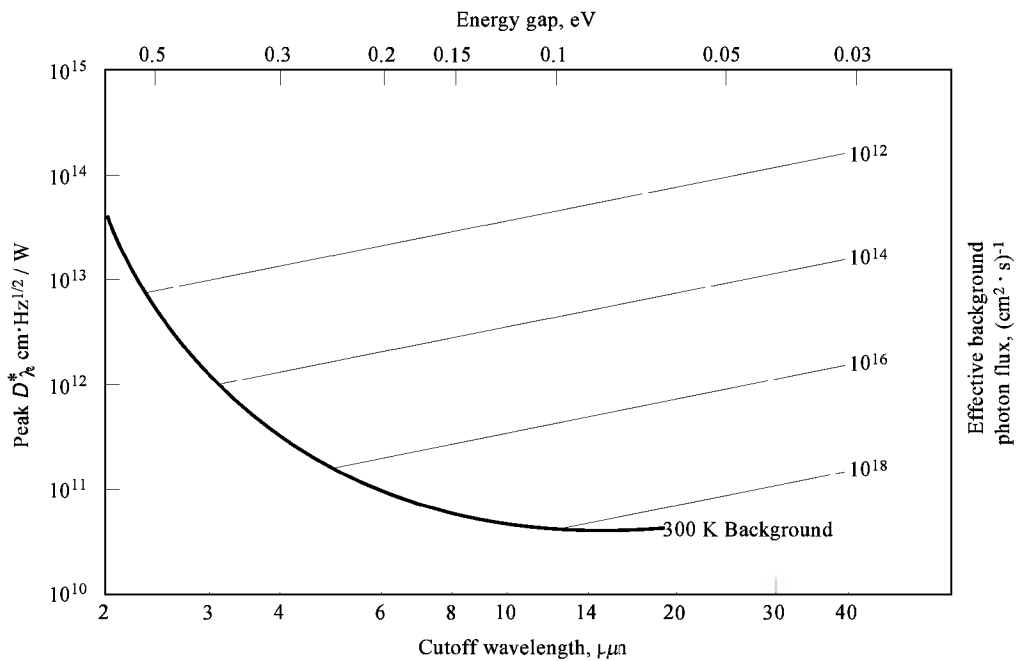


Fig. 3. Ideal photon detector sensitivity as a function of cutoff wavelength. Lower background flux generates less photon-induced noise giving higher sensitivity. The sensitivity limit for the condition of 300 K background temperature and hemispherical (2π) field of view is shown.

The cooling needed to achieve a given detectivity for intrinsic-type detectors may be readily calculated based on the temperature dependence of the minority carrier lifetime and mobility. The calculated values for indium arsenide [1303-11-3], InAs [1303-11-3], and two compositions of HgCdTe, are given in Figure 4. Experimental data for lead sulfide [1314-87-0], PbS [1314-87-0], and three HgCdTe detectors having different compositions, are also shown. The importance of cooling is obvious.

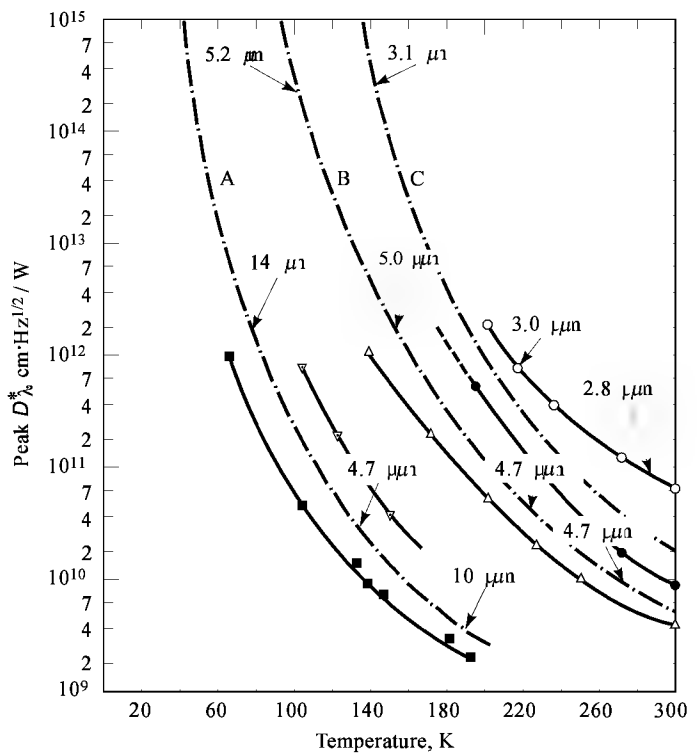


Fig. 4. Sensitivity as a function of detector temperature showing (—) experimental results for HgCdTe; (■ and △) $\text{Hg}_{1-x}\text{CdTe}$, $x = 0.185$ and $x = 0.280$, respectively; (▽) InSb; (●) InAs; and (○) PbS. Also shown are theoretical values (---) based on thermal equilibrium statistics, A and B, $\text{Hg}_{1-x}\text{CdTe}$ where $x = 0.195$ and $x = 0.29$, respectively; and C, InAs.

Detector cooling often is accomplished by providing good thermal conductivity to a suitable cryogen (2). The most readily available coolants are solid carbon dioxide [124-38-9], CO_2 [124-38-9], at 195 K, liquid nitrogen, N_2 , at 77 K, and liquid helium, He, at 4.2 K (see CARBON DIOXIDE; HELIUM GROUP; NITROGEN). If the desired temperature is not below ca 200 K, a thermoelectric cooler may be used. Joule-Thompson and expansion engine coolers are available for operation to 4.2 K and below. A cooled detector must be contained in a vacuum vessel which thermally isolates the detector element from other than the coolant by low thermal conduction materials and vacuum (see VACUUM TECHNOLOGY). If low absorbing extrinsic detectors are used, it is customary to enclose the detector element in a cooled chamber with reflecting walls, ie, in an integrating chamber, in order to obtain multiple passes of the radiation. The latter also can be accomplished by shaping the element so as to ensure multiple, total, internal reflections. Thin detectors can be sized to produce optical resonance in the detector element making the detector itself an optical resonant cavity.

The choice of a detector for a specific application should be made in order to minimize the cooling requirements and the magnitude of the background radiation noise; therefore, in detector selection the cutoff wavelength should be only slightly greater than that required by the application. If the

signal radiation has a substantial path length through the atmosphere, then atmospheric transmission characteristics must be considered. Because of the presence of CO₂ and H₂O vapor, many absorption bands appear in the atmospheric transmission spectrum. Windows of high transmittance occur between the bands. The principal windows are at 1.0–2.5, 2.9–4.2, 4.4–5.3, 7.5–14.0, and 16–23 μm (2).

Photodetector Modes of Operation

The need for detectors with high performance and low cost has resulted in fewer than two dozen types of detectors which are available as commercial products. These are made from only 10 basic semiconductor elements or compounds. Popular detectors, their modes of operation, cutoff wavelength, temperatures of operation, response times, responsivities, and detectivities are listed in Table 1 (15). The values chosen are for near-optimum performance at a given temperature. The development of the silicon charge-coupled device (CCD) detector has been an important consequence of the microprocessor industry. Silicon photovoltaic and charge-coupled detectors operate near room temperature. Cadmium sulfide [1306-23-6] and germanium detectors also require no cooling. Wide acceptance of the ternary compound semiconductor mercury–cadmium–telluride [29870-72-2] (MCT), HgCdTe, began in 1972 as a photoconductor. Requirements for large focal planes has since led to significant developments of the MCT photodiode focal plane (16). The controllable (by composition) band gap of MCT in each case results in a cutoff wavelength that is tailored for a specific application (see Fig. 4). Lead sulfide, indium arsenide, and MCT (40% at. wt CdTe) detectors can be operated at 300 K but perform much better at lower temperatures (2,14).

Four principal modes of semiconductor-based photodetectors are the charge-coupled device (CCD), photoconductor, photodiode, and bolometer. The Schottky barrier mode using internal photoemission has been well developed using platinum silicide, but low quantum efficiency restricts the range of uses. The quantum well infrared photodetector (QWIP) is included herein because artificially structured materials such as this superlattice detector are representative of a new metallurgy. The semiconductor materials for most applications are Si, Ge, GaAsP, InSb, PbS, CdS, and HgCdTe. These materials grown as single crystals and thin polycrystalline or amorphous films are doped with various elements such as boron, phosphorus, indium, gold, etc, to control the polarity of the conductivity and for selective optical absorption by the introduction of impurity states in the forbidden energy gap. Photodetection using these materials extends from the ultraviolet (0.3 μm) to long wavelength infrared (200 μm). Higher sensitivity, especially in the infrared, can be achieved by cooling the detector and its immediate surroundings to low temperature, typically 77 K. Interfacing with signal processing electronics can be difficult when the photon signal is weak and imaging systems require sophisticated signal processing to handle the hundreds of thousands of pixels on a focal plane.

Charge Mode Detector. Charge mode devices (17,18) are a group of detectors which utilize the metal–insulator–semiconductor (MIS) capacitor as their basic constituent. An MIS capacitor is a device consisting of a nondegenerately doped semiconductor substrate, a thin dielectric layer, and a metal gate, as shown in Figure 5 (see THIN FILMS). The application of a potential difference across the metal gate and the semiconductor produces a surface potential barrier and an associated depletion region in the semiconductor. Periodic pulsing of the gate-to-substrate bias creates a nonequilibrium condition suitable for the storage of optically generated free carriers.

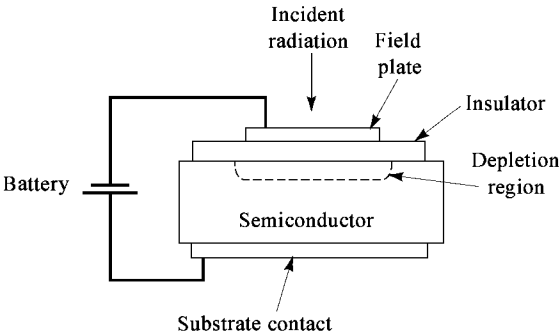


Fig. 5. Device model for an MIS photodiode, the basic building block of the CCD photodetector. The depletion region is generated by the battery potential. See text.

The model for an MIS capacitor is shown in Figure 6 and includes a metal gate-to-semiconductor bias of V_g volts producing an insulator bias, V_i , and surface potential, Ψ_s . The negative for n -type material gate charge, Q_g , produces an ionic space charge, Q_d , and inversion layer charge, Q_p . From the basic relationships (19) that follow,

$$V_g = V_i + \Psi_s \tag{13}$$

$$Q_s = Q_d + Q_p \tag{14}$$

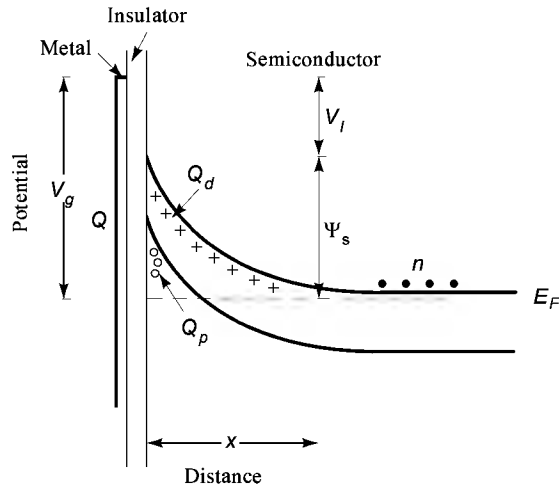


Fig. 6. Band model for the charge mode detector biased to deep depletion. The charge, Q_p , integrates in the potential well defined by the insulator and valence band. See text.

where $Q_s = C_I V_I$ and $Q_d = C_d \Psi_s$ then

$$\Psi_s = V_g (1 + C_d / C_I)$$

(15)

where C_d is the capacitance associated with the space-charge layer geometry of charge n at distance x . From semiconductor theory,

$$\Psi_s = n q x^2 / 2 \epsilon$$

(16)

where ϵ is the dielectric permittivity of the semiconductor and, from geometry,

$$C_d = \epsilon A / x$$

(17)

The basic MIS condition is obtained by combining equations 13–17:

$$\Psi_s + (\epsilon q n / 2)^{1/2} A (\Psi_s)^{1/2} / C_I = V - Q_p / C_I$$

(18)

At initial turn-on, the inversion charge $Q_p = 0$ and when n is small and C_I is large, $\Psi_s \approx V$, that is, a potential well of nearly V is formed. Thus, the charge-handling capacity, Q_{FW} is

$$Q_{FW} \approx C_I V$$

(19)

and can be seen to be dependent upon both the physical dimensions of the capacitor and the applied voltage. Typical full well capacities range from 20,000 carriers to greater than 2,000,000 carriers, depending on the specific charge mode device.

As time progresses, charge Q_p is generated by photon currents, semiconductor dark currents, diffusion, depletion, surface, avalanche, and tunneling. The time required to fill the well by dark currents is the storage time, τ_{st} given by the following.

$$\tau_{st} = C_I V / J_d$$

(20)

The storage time is the maximum time the potential well can be used for collection of photon-generated charge. Because J_d depends strongly on band gap energy and temperature of operation, τ_{st} in practice, extends from hours for silicon at 300 K to a few microseconds for HgCdTe, where the cutoff wavelength, $\lambda = 12 \mu\text{m}$ at 77 K. Indium antimonide operates in a few tenths of a second at 77 K (20).

As previously indicated, the MIS capacitor can be utilized as a photon detector. A negative pulse on an n -type semiconductor generates a deep depletion region at the surface which, in turn, acts as a storage site for photogenerated holes. This type of detection is true charge integration and is the simplest charge mode detection concept. Other MIS devices exist which contain arrays of overlapping, ie, metal gates in a given detector element. These MIS-based detectors are called charge-transfer devices (CTD). The two most common forms of CTDs are charge-injection devices (CID) and charge-coupled devices (CCD).

The CID is a dual gate version of a CTD. In the CID, two gate electrodes form two distinct but physically adjacent depletion capacitances, as shown in Figure 7a. Under normal operating conditions each of the two gate electrodes is biased to form a depletion region in the underlying semiconductor. However, the surface potential minimum under one of the electrodes is much greater than that of the other electrode. As a result, the photogenerated charge produced in the detector collects in the deeper potential minimum. To measure the number of stored carriers, the potential difference between the collection electrode and the semiconductor substrate is reduced, thereby collapsing the storage well and forcing the collected charge to transfer to the region controlled by the second gate electrode. The collected charge can, in principle, be quantified by monitoring the change in surface potential of this second electrode during the transfer operation. In practice, this is accomplished by monitoring the instantaneous current through the electrode at the instant of charge transfer. The detector is reset to the empty state by simultaneously collapsing the depletion regions of both capacitors. This action forces the free carriers into the semiconductor substrate where they are subject to recombination mechanisms. Thus a key feature of a CID is that photogenerated charge is both collected and sensed inside a single detector element. Signal charge is transferred from that element simply as a means to eliminate it from the detector prior to the next detection cycle.

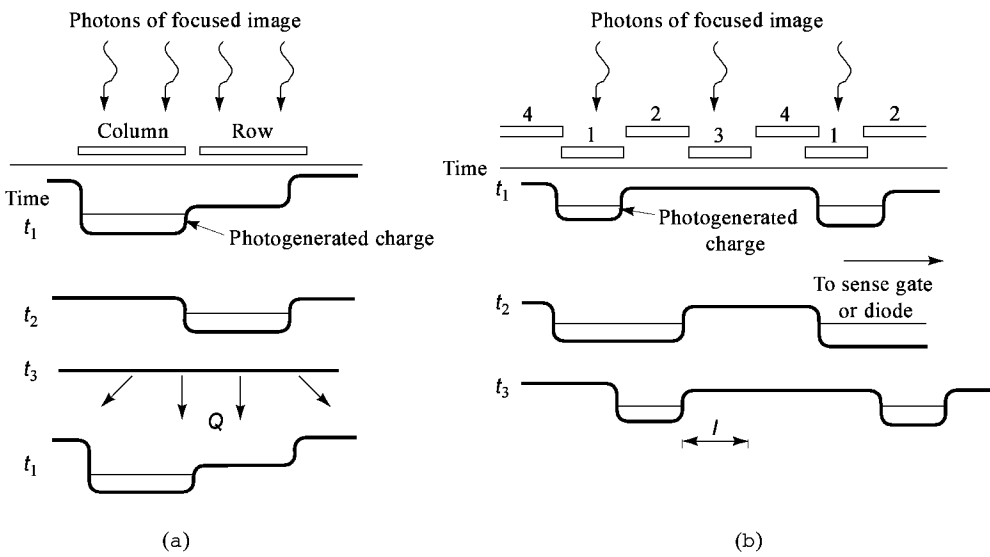


Fig. 7. Conceptual cross-section views of (a) a CID imager showing charge integration (t_1), charge measurement (t_2), and charge removal (t_3); and (b) a CCD shift register showing charge transfer from a phase one well to a phase two well.

The charge-coupled device (CCD) is a multigate CTD. As for the CID, the CCD gates are physically adjacent and are periodically pulsed, inducing the formation of deep depletion regions in the underlying semiconductor. However, unlike the CID, the CCD physically transfers the collected photogenerated charge from a given detector site to a separate charge measurement structure. Figure 7b shows a conceptual cross-section view of a CCD shift register. As the photons of an image which is focused on the device generate charge, the MIS gates of the shift register are biased in sequence to

produce, continually, a shifting potential well pattern. Thus, at time t_1 , phase one is on and charge is integrating in the potential well beneath the MIS phase one gates. At time t_2 , phases one and two are turned on and the charge distributes beneath phase one and phase two gates. At time t_3 , phase one is turned off and the charge is beneath phase two only. Pulsing phases three and four in the proper sequence with phase one and two moves the charge packet through the shift register. For an area array CCD the imaging portion of the array transfers charge, once per line interval, into a readout shift register in a manner equivalent to that described for the imaging shift register. Each charge packet in the imaging register is then shifted in a discrete fashion to the charge-detection structure residing at the end of the register. This sense electrode or diode is in turn typically connected to a low noise charge-sensitive amplifier which converts the signal charge to a voltage. After all charge packets in the shift register have been read out, the next line of information from the imaging portion of the array is transferred and the process is repeated.

Photoconductors. The photoconductor (21) is a semiconductor resistor that lowers in resistance when photons generate excess carriers. The two general classes of photoconductor are extrinsic, which is characterized by impurity states that are emptied by photons, and intrinsic, where the photons excite electrons directly from the valence band to the conduction band (Fig. 1). The photoconductor is connected in series with a constant load resistor and current source. The ambient light or background radiation sets up a steady-state resistance in the detector. Changes in illumination produce changes in resistance and thereby changes in the voltage across the load resistor. When the load resistance is much larger than the detector resistance, signal voltage is given by equation 21:

$$V_S = \frac{\eta \phi_s \tau V}{n t} \tag{21}$$

where the signal flux, ϕ_s , is caused by the changing background. At low bias the signal is linear with the bias voltage, V , but at high bias the minority carrier sweep time becomes less than the lifetime, τ , and the signal saturates. The majority carrier density, n , is typically 1×10^{15} to $1 \times 10^{16} \text{ cm}^{-3}$. The thickness, t , is typically one to three absorption lengths (8–20 μm).

Photodiodes. Photovoltaic detectors (21) generate a voltage (open circuit) or a current (closed or integrating circuit) when receiving radiation and therefore do not require an external bias for operation. The charge integrated on a capacitor of the integrating circuit is proportional to the intensity of radiation. The current generated by the photovoltaic cell, I_p , is given by the following.

$$I_s = \eta \phi_s q A \tag{22}$$

The quantum efficiency can be determined using the measured current and the aid of a calibrated photon source, eg, a blackbody. The detector area A should be measured by a spot-scanning apparatus, because the minority carrier diffusion length effectively increases the diode photon-active area. A photovoltage can arise from the presence of internal fields which are formed at the interface between a semiconductor and a metal (Schottky) barrier, at the interface between different dopings in a semiconductor (homojunction), and at the interface between two different semiconductors (heterojunction). Light releases mobile electric carriers within these barrier layers. These carriers are separated by the field and produce a photovoltage between them, as indicated in Figure 8. Photovoltages can be produced in the bulk of a semiconductor because of the difference in the diffusion of electrons and holes or if there is a gradient in the impurity concentration within the material, but these mechanisms are not in wide use.

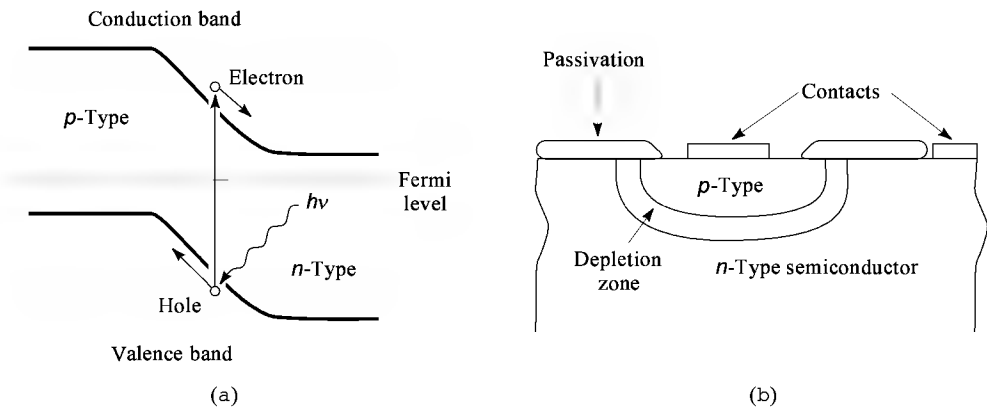


Fig. 8. The photodiode detector: (a) band model where the photon generates electron–hole pairs that are separated by the built-in potential setting up a photocurrent; (b) physical model for a planar diode. The passivation is typically SiO₂ for Si diodes, an In oxide for InSb diodes, and CdTe for HgCdTe diodes.

The impedance area product, ZA , of a diode may be obtained from differentiation of the diode current–voltage relationship and may be expressed as follows, at zero applied bias.

$$ZA = \frac{kT}{BqJ_d} \tag{23}$$

Under little or no illumination, J_d must be minimized for optimum performance. The factor B is 1.0 for pure diffusion current and approaches 2.0 as depletion and surface-mode currents become important. Generally, high crystal quality for long minority carrier lifetime and low surface-state density reduce the dark current density which is the sum of the diffusion, depletion, tunneling, and surface currents. The ZA product is typically measured at zero bias and is expressed as $R_{\phi}A$. The ideal photodiode noise current can be expressed as follows:

$$i_N^2 = 2qA(2J_d + J_{\phi}) \Delta f \tag{24}$$

where J_{ϕ} is the photon current density. Equations 23 and 24 can be used to determine the cooling requirements for photodiode detector performance since diffusion current depends exponentially on temperature (22). Because the diffusion current represents a valence band-to-conduction band excitation, the thermal limit, when $J_d > J_{\phi}$, is essentially the photoconductive limit (see Fig. 4). Combining equations 5, 8, 14, 23, and 24 gives the following:

$$D^* = q\eta \ h\nu(4kT \ ZA + 2qJ_{\phi})^{-1/2} \tag{25}$$

Thus, ZA must be increased to achieve BLIP D^* such that $ZA > 2 \ kT \ qJ_{\phi}$. When $J_{\phi} > J_d$, equation 25 reduces to the ideal photon limit of equation 12. The response time is equal to the ZC product, where C is the diode capacitance, or when using an integrating amplifier, the response time is determined by the closed loop gain.

Bolometers. The bolometer (23) has made a comeback as a popular detector thanks to advances in micromachining technology. When applied

to silicon it is feasible to fabricate large arrays of bolometers and detecting elements having very small thermal mass. The imaged infrared radiant power slightly heats the bolometer film a few millidegrees kelvin causing a lowering of the electrical resistance. The resulting change of the bias current is the signal. The noise components are 1 f, Johnson-Nyquist, thermal conductance, and photon. The photon noise is actually the radiative part of the thermal conductance noise.

From bolometer theory (1) the change in film temperature is proportional to the absorbed power and thermal resistance (inverse of the thermal conductance) and is given by the following:

$$\Delta T_b = J_s A \epsilon_a R_{th} \frac{\left(1 - \exp\left(\frac{\tau_s}{R_{th} C_h}\right)\right)}{\left(1 + \exp\left(\frac{\tau_s}{R_{th} C_h}\right)\right)} = J_s A \epsilon_a R_{th} \tanh\left(\frac{\tau_s}{2 R_{th} C_h}\right) \tag{26}$$

where J_s is the differential power density at the detector caused by a change in target temperature in W cm^{-2} ; A is the active pixel area, $1.2 \times 10^5 \text{ cm}^2$; ϵ_a is the ir absorption efficiency, 90%; R_{th} is the thermal resistance and is approximately $1.5 \times 10^7 \text{ K W}$; τ_s is the half period of chopped radiation ($1/2f$), typically 16.7 ms (30 Hz modulation); and C_h is the heat capacity of the pixel element and is approximately $0.7 \times 10^{-9} \text{ J K}$.

The power density, ΔJ , may be calculated using Plank's radiation law. For a 300 K scene temperature and the spectral region from 8 to 12 μm ,

$$J_s = 2.0 E^{-4} \left(\epsilon_0 \Delta T_s \frac{\Omega}{\pi} \right) \tag{27}$$

where ϵ_0 , the optical efficiency, is typically 80%, ΔT_s is the differential target temperature, and Ω is the collection solid angle of the optical system ($= \pi \sin^2(\tan^{-1} 1/2f_n)$).

For the bolometer detector having the above characteristics and operating with $f = 1$ optics ($f_n = 1$) the peak to peak temperature change of the α -Si film is approximately 3 mK per one degree kelvin change in scene temperature.

Signal. The peak to peak signal voltage, V_s , is proportional to the bias power and can be expressed as follows:

$$V_s = \alpha (P R_b)^{1/2} \Delta T_b \tag{28}$$

where α is the temperature coefficient of resistance and is 2.8° K for α -Si, P is the bias power for the bolometer element (typically 0.3 μW), and R_b is the bolometer resistance (typically $3 \times 10^7 \text{ }\Omega$ for α -Si and $1 \times 10^4 \text{ }\Omega$ for VO).

The voltage responsivity is simply given by equation 29:

$$R_V = \frac{V_s}{J_s A} = \alpha (P R_b)^{1/2} \epsilon_a R_{th} \tanh\left(\frac{\tau_s}{2 R_{th} C_h}\right) \tag{29}$$

Responsivity values are typically $5 \times 10^5 \text{ V W}$ range for nominal bias.

Noise. The noise components may be expressed as voltage. Excess 1 f noise:

$$V_f = \int_{\Delta f} \left(\frac{\gamma}{f^{1/2}} \right)^2 df = K_V (P R_b)^{1/2} \tag{30}$$

where K_V is the 1 f noise coefficient and is typically in the range from 6 to 12 μV per volt bias when integrated from 1 to 100 Hz. Johnson-Nyquist noise, V_{JN} , is

$$V_{JN} = (4kT R_b \Delta f)^{1/2} \tag{31}$$

where k is the Boltzmann constant, T is the detector temperature (300 K), and Δf is the noise bandwidth (100 Hz). Thermal conductance noise, V_{TC} , is

$$V_{TC} = \alpha (P R_b)^{1/2} \left(\frac{kT^2}{C_h} \right)^{1/2} \tag{32}$$

when the thermal response time $R_{th} C_h$ is less than the frame time.

NEP. The noise components may be combined with the responsivity to give the noise equivalent power (NEP) equation:

$$\text{NEP} = \frac{(V_{JN}^2 + V_{EX}^2 + V_{TC}^2)^{1/2}}{R_V} \tag{33}$$

Measured values are typically less than 100 pW for a 30-Hz scene modulation.

Detector Fabrication and Performance

Significant improvements in detector performance and focal plane array development have occurred since the early 1980s (24,25). The spectral sensitivities of several photodetectors are given in Figure 9 covering the spectrum from ultraviolet to far infrared. The typical operating temperature is given and the theoretical limit is that calculated for the detector exposed to hemispherical radiation from a 300 K background.

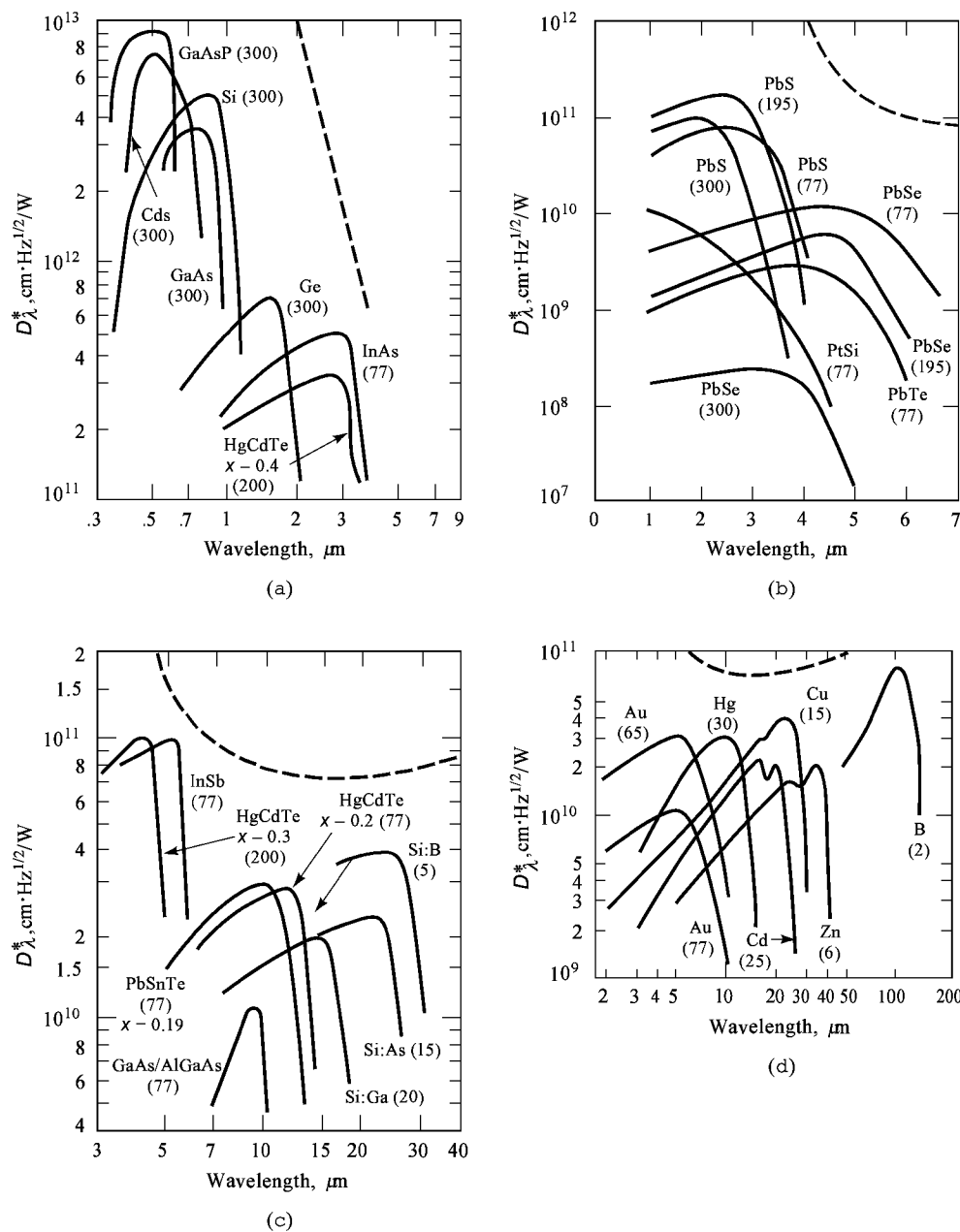


Fig. 9. Spectral sensitivity of detectors where the detector temperatures in K are in parentheses, and the dashed line represents the theoretical limit at 300 K for a 180° field of view. (a) Detectors from near uv to short wavelength infrared; (b) lead salt family of detectors and platinum silicide; (c) detectors used for detection in the mid- and long wavelength infrared. The Hg_{1-x}CdTe, InSb, and PbSnTe operate intrinsically, the doped silicon is photoconductive, and the GaAs/AlGaAs is a structured superlattice; and (d) extrinsic germanium detectors showing the six most popular dopants.

Charge-Coupled Devices and Imaging Arrays. The single most popular type of visible photon detector is the charge-coupled device (CCD). Developed in 1970 (26,27) as a dynamic memory storage device, the CCD was quickly adapted for use as an imaging device. The CCD has been a resounding commercial success, becoming the detector of choice in the video camera market, and demonstrating great potential as the optimum detector for future electronic still photography (qv) systems. Furthermore, since its inception the CCD has also been an invaluable tool in a wide range of scientific applications involving the detection of near-infrared, visible, ultraviolet, and x-ray photons, as well as in the detection of ionizing radiation and charged particles.

CCDs are sensitive photon detectors. The physical mechanism governing photon detection in silicon depends on the energy of the incident photons. For the near-infrared to visible spectral band, silicon performs as an indirect band gap semiconductor. Incident photons are converted to electron-hole pairs through interactions with the silicon lattice, producing one electron-hole pair for each absorbed photon. For higher energy photons such as those of the extreme uv and x-ray spectral bands, direct band gap detection dominates. In this case, one electron-hole pair is generated for every 3.65 eV of energy contained in the absorbed photon. If the scene of interest produces only a small number of high energy photons impinging on the detector array during each frame interval, the number of signal electrons in each pixel is a measure of the energy of the incident photon. In such applications, CCDs can be used to obtain spectral as well as spatial information (28).

Utilization. CCDs are often grouped into two broad categories: commercial-grade and scientific-grade. For both categories the format and required performance of the CCD are driven by the application for which the array is designed. In general, commercial-grade CCDs are designed to optimize image resolution and color fidelity. On the other hand, scientific-grade CCDs are designed to maximize photon quantum efficiency and minimize device noise. The vast majority of available CCDs have been designed for commercial applications. By virtue of the relatively small number of devices that are supplied yearly for scientific applications, and the more stringent performance requirements, scientific-grade CCDs are typically more expensive than their commercial counterparts. Commercial-grade CCDs can be adapted for use in scientific applications where a moderate photon level is present, such as some forms of visible spectroscopy.

Most CCDs are specifically designed for video camera applications, which detect photons in the visible portion of the electromagnetic spectra. Video camera CCDs have specific camera formats compatible with standard video display systems. In the United States, video displays utilize the RS-170 standard, commonly known as the National Television System Committee (NTSC) 525-line television standard. The RS-170 standard places many constraints on imaging devices. These devices must have a frame time of 1/30 second, with two interlaced image fields of 1/60 second each required during each frame. The standard luminance electrical bandwidth of 4.2 MHz coupled with the 4:3 picture aspect ratio require the pixel count and the pixel size of the commercial CCD array to fit within narrowly defined regimes. Specifically, to meet the electrical bandwidth requirements of the RS-170 standard, monochrome imagers must have a minimum of 240 pixels in the vertical dimension and a minimum of 450 pixels in the horizontal dimension. The distance along the diagonal of the active imaging portion of the CCD must be one of several specific values in the 5.46–25 mm range. The specific

value depends on the particular camera format, eg, 8 mm, VHS, etc, for which it is designed. These format and size requirements dictate a rectangular rather than square CCD pixel design. Data rates from the imaging devices must be chosen so that the image scene readout time matches the RS-170 interval. The minimum data rate for monochrome imagers is 6 MHz. The actual rate spans from 6 to 10 MHz, depending on the specific device.

Color applications require more complex image detection schemes owing to the need for spectral differentiation. Signal samples from at least three distinct spectral bands are required to accurately reproduce a color image. Typically a monolithic color filter array consisting of alternating windows of appropriately colored filter media is attached to the CCD to accomplish color differentiation. A common approach to obtain the scene information required for faithful color image reproduction is to increase the horizontal pixel count of the CCD by a factor of three (29). In this method, a filter with alternating red, blue, and green stripes is then attached so that each pixel column aligns with a filter stripe. This color CCD utilizes three adjacent pixels, one red, one green, and one blue, to represent a single spatial sampling of the scene. In effect, the device can be thought of as three integrated monochrome CCDs. The color CCD has three on-chip output channels, one for each spectral component. As each pixel is read, the simultaneously generated signal information from each channel is remixed in a fashion that correctly simulates a color display. More ambitious methods exist for achieving color image reproduction (30,31). These methods utilize the horizontal pixel count found in monochrome CCDs but require intricate clock control methods to appropriately intermix the spectral components for display during the actual integration interval. The intermixing is produced by the use of a repetitive mosaic pattern of colored filters, typically cyan, yellow, and green. The imaging results from these devices rival those obtained from the three-channel method.

Video camera applications obligate the CCD imager to have an internal, electronic, image shuttering capability in order to clearly differentiate the data in one image frame from that of the next. The two most common forms of internal shuttering are frame transfer and interline transfer (32). In the frame transfer technique, the image gathered from one data frame is rapidly transferred from the contiguous image collection region to a secondary, optically shielded storage site that resides between the image collection and shift register regions. The image held in the shielded region is then read out during the time interval of the next image formation in the optically active area. In the interline transfer technique, each pixel in the imaging portion of the array is segmented into an optically active portion and a shift register portion. The image gathered from one data frame is quickly transferred from the optically active to the shift register portion of the pixel, and read out during the next image collection interval. Unlike the frame transfer method, no additional image storage section is required for the interline transfer method. However, the frame transfer device is more efficient at photon detection, since there are no optically shielded features in the image collection region.

The requirement of internal shuttering forces the CCD to be illuminated from the front side. Photons impinging upon the front surface of the detector must pass unimpeded through the gate electrode layers in order to be collected in the photon-sensitive silicon beneath. Typically some loss does occur for photons of wavelengths shorter than about 550 nm due to absorption in the gate electrodes. For standard polysilicon gate CCDs, photons of wavelength less than 400 nm are absorbed by the gate material. This absorption process affects photons spanning the uv to low energy x-ray spectrum. Photons of wavelengths greater than 650 nm pass easily through the gate electrodes, but owing to the lower value of the absorption cross section of silicon, a several micrometer path length is required to provide a high probability of absorption. Deeply generated signal charge can produce image smearing due to charge diffusion prior to collection in the potential well of the CCD pixel. Both the photon collection efficiency and image quality are strong functions of the particular CCD design. Typically video CCDs provide sensitivities spanning the 50–500 mA⁻¹ W range and have minimum detectable signal levels of from 40 to 400 photons at the standard 16 ms image integration interval.

CCDs are being actively developed for use in the field of electronic still photography to provide a means of electronically gathering high quality images. The ideal CCD-based still camera would gather an image of comparable quality to that of a 35-mm film camera. However, unlike film-based cameras, the image would be stored on electronic media such as a floppy or optical disk (see INFORMATION STORAGE SYSTEMS). The image could then be read into a computer-based system, edited or enhanced as desired, and stored again onto the digital media. Hard copies would be obtained on photographic film by a separate image transfer system. Still photography places more stringent specifications on a CCD than those required of RS-170 video cameras. Still-photography CCDs must have very high resolution. To match the image quality of 35-mm film, a CCD with a format of greater than 4000 by 4000 pixels is required. For optimum image reproduction the pixels should be square. The device should operate at low light levels and be of a size to operate using standard 35-mm optics.

The development of high quality still photography is still in its infancy. Devices as large as 5120 by 5120 pixels have been built and demonstrated as high quality monochrome imagers (33); however, the 87-mm diagonal size makes such a device too large to be useful in commercial photographic equipment. Further effort is required to address the issues of color imaging and to decrease the size of the CCD pixel without adversely affecting the dynamic range of the device. The advances should also serve to further the development of imagers for high definition television (HDTV).

Whereas commercial video CCDs can sometimes be adapted for scientific endeavors, these devices have format and operability constraints which negatively impact their performance in many applications. Scientific CCDs are somewhat more difficult to categorize because of the wide range of capabilities available in such devices. The performance features of a specific scientific CCD are determined by the device design. A few of the more significant performance features which may be found in a given scientific CCD include the following: (1) spectral response extending over a wide range of photon and charged particle energies; (2) noise levels as low as one electron per pixel, which result in very high signal-to-noise ratios; (3) very high charge-transfer efficiencies to achieve no measurable degradation in spatial or spectral information upon image readout; (4) dark current generation rates permitting several minute integration times at room temperature and multiple hour integration times at cryogenic temperatures; (5) high photon detection efficiency; and (6) high frame rate operability. In addition, scientific CCDs are typically operated in the full-frame readout mode (32), in which an external mechanical shutter shields the imager during scene readout.

Scientific CCDs have been used in a wide range of applications. In the visible and uv portions of the spectrum, scientific CCDs have proven to be effective tools for low photon flux applications. Specifically, CCDs have been utilized as image detectors in ground-based telescopes as well as in space-borne systems such as the Wide-Field Planetary camera of the earth-orbiting space telescope (34), the Giotto mission to Halley's comet (35), and the Galileo mission to Jupiter (36). Figure 10 is a photomicrograph of the 800 × 800 pixel scientific CCD designed specifically for the Galileo mission. The photomicrograph was taken from a complete Si wafer after device processing but prior to device packaging.

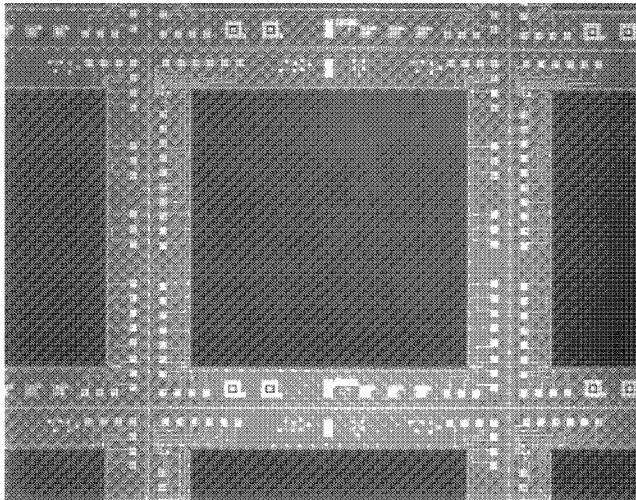


Fig. 10. The 800×800 pixel scientific CCD specifically designed for the Galileo mission to Jupiter. The devices are shown in wafer form after initial performance screening measurements.

The same features that make scientific CCDs excellent devices for astronomy, that is, high photon collection efficient and low readout noise, also make CCDs excellent tools for chemical analysis (37). CCDs can be utilized in many forms of spectroscopy (qv), including absorption (38), fluorescence (39), luminescence (40), emission (41,42) and Raman (43), over a spectral range of 0.1 to 1100 nm. The high dynamic range (typically >100 dB) of the CCD is an invaluable property for identifying weak spectral lines. Furthermore, the inherent integration of many highly sensitive photodetector elements in a small area has, in some instances, allowed revolutionary experimental designs (44). Several visible light spectrometer systems which utilize CCDs as the imaging device are commercially available.

In the x-ray portion of the spectrum, scientific CCDs have been utilized as imaging spectrometers for astronomical mapping of the sun (45), galactic diffuse x-ray background (46), and other x-ray sources. Additionally, scientific CCDs designed for x-ray detection are also used in the fields of x-ray diffraction, materials analysis, medicine, and dentistry. CCD focal planes designed for infrared photon detection have also been demonstrated in InSb (47) and HgCdTe (48) but are not available commercially.

Fabrication. Although CCDs have been fabricated in many semiconducting materials such as Ge (49), InP (50), and HgCdTe (51), by far the most readily available devices are those which utilize Si as the semiconductor. There are several common types of silicon CCDs. All share certain processing steps, and most utilize *p*-type silicon as the semiconducting material. Silicon single crystals are grown (52) in conventional Czochralski vertical pullers using a single-crystal silicon seed dipped and rotated in a silicon melt. The large boules of silicon are sawed into wafers. The following fabrication discussion describes a process for the creation of a generic four-phase CCD in *p*-type silicon (53).

A cross-section view of a four-phase Si-based CCD is shown in Figure 11. The starting wafer is a thin *p*-type silicon layer grown epitaxially on a degenerately doped *p*⁺ silicon substrate of roughly 500 μm thickness. The epitaxial layer is typically 8–15 μm in thickness and is doped with boron to a nominal resistivity of 10 ohm-cm. The initial step of the fabrication process consists of a cleaning procedure designed to ensure a nearly defect-free surface. Next, a thin layer of protective SiO_2 is grown on the silicon surface by the use of elevated temperatures in a steam and oxygen ambient. Channel stops, thin stripes which provide lateral containment of the stored charge, are created by high energy implantation of *p*-type ions into appropriate regions of the silicon. Most CCDs are designed with a thin layer of *n*-type dopant at the silicon surface which serves to hold the stored charge physically away from the silicon surface. This dopant layer, called the buried channel, is produced by implanting *n*-type dopant ions such as P or As into the silicon. After the implantation of the channel stops and buried channel, the original SiO_2 layer is removed and a fresh, undamaged layer of SiO_2 is regrown. Next a layer of heavily impurity-doped polysilicon is deposited and patterned to form the ϕ_2 and ϕ_4 gate electrodes. An additional layer of SiO_2 is then grown to provide an effective insulating layer atop the first polysilicon layer. The ϕ_1 and ϕ_3 gate electrodes are formed from a second layer of heavily impurity-doped polysilicon. A low resistance material such as aluminum is then used to form electrical connections between appropriate parts of the CCD. Typically this same metal layer is utilized to form bond pads which connect the CCD to external control signals from off-chip electronics. However, some process sequences require the inclusion of a second metal layer for this purpose. Finally the CCD is covered with an overcoat of protective material such as SiO_2 or borophosphosilicate glass. This 500–1000-Å thick layer serves as a barrier between the environment and the contamination-sensitive CCD.

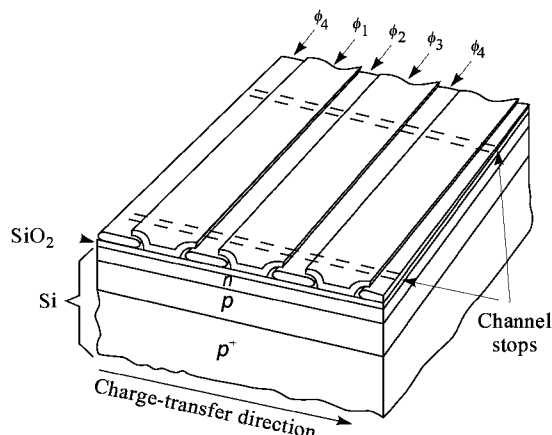


Fig. 11. Cutaway view of a CCD shift register where the ϕ_i represent gate electrodes. Voltage pulses applied to the phase gates move photogenerated charge in the charge-transfer direction. The channel stops confine the charge during integration and transfer. See text.

Device Type vs Application. The application for which the CCD is designed dictates the variants to the process that are utilized to provide the desired performance enhancements. Consider as an example the effect of front-side illumination of the CCD on photon collection efficiency.

Applications which require very high photon collection efficiency in the visible blue, extreme uv, or x-ray spectral bands are not well served by front-side illumination of the CCD. A useful CCD variant for such applications is the back-side illuminated CCD. Back-side illuminated CCDs undergo additional processing to remove the underlying *p*⁺ substrate from the *p*-type epitaxial layer. As the name implies, photons impinge upon the device from the back side, thereby avoiding the absorption layers of gate electrodes present in front-side illuminated devices. In this manner the photon collection efficiency of the device can be improved in the blue, uv, and low energy x-ray photon regimes. Substrate removal is accomplished through the use of an etchant with an etch rate that is highly dependent upon the concentration of boron dopant found in the silicon. The process of substrate removal is well understood and highly repeatable. However, thinned (to ~25 μm) back-side illuminated CCDs are typically more costly than thick front-side illuminated devices. The additional cost is associated with the reduction of mechanical rigidity produced by the removal of the substrate.

Other CCDs have special processing steps that lower the rate at which surface dark current is generated during the interval in which signal charge is collected. One such device is known as the virtual phase CCD (VPCCD) (54). The VPCCD utilizes a series of implanted dopant layers near the surface of the silicon as a means of removing one of the polysilicon deposition steps required for electrode formation. The implanted layers serve as virtual electrodes which force the surface potential of the silicon in these regions to a constant bias during operation. These virtual electrodes replace two of the four gate phases described in the generic CCD process. The two remaining gate control phases are incorporated into a single physical gate electrode, again by means of ion-implanted layers in the silicon. Thus, the top side of the VPCCD pixel is only partially covered by the polysilicon gate electrode. The ion-implanted layers provide storage capability and unidirectional charge transfer as the single polysilicon electrode phase is clocked. The virtual electrodes also serve to greatly reduce the rate of surface-related dark current generated in the CCD. At 25°C the nominal dark current value of a VPCCD is 0.5 nA/cm², which is nearly 20 times lower than the dark current found in a typical buried channel CCD.

A second device having additional specialized ion-implanted layers is the multipinned phase (MPP) CCD (55). The MPP device is a compromise between a multiphase and a virtual phase CCD, and is becoming the CCD of choice for scientific applications. As in the generic multiphase CCD, the entire charge storage region is covered by polysilicon gates. However, prior to the electrode deposition, additional dopant layers are ion implanted into the silicon. These implanted dopant layers enable the MPP CCD to integrate charge with the applied gate bias set so as to attract opposite polarity charges to the Si-SiO₂ interface. This method of operation produces the lowest possible dark currents, rivaling or exceeding the performance of the virtual phase

CCD. Although front-side illuminated devices are available, the MPP CCD is typically back-side illuminated in order to achieve state-of-the-art photon collection performance. Even with this stipulation the MPP CCD is still a popular detector due to its availability in various array sizes and formats specifically designed for scientific applications.

Silicon Photodiodes. The popularity of silicon photodiodes is directly related to the ability to detect photons over a spectral range spanning the near-infrared to low energy x-ray regimes. Typical spectral response characteristics are given in Figure 9a. The fast response time of less than 1 μ s is attractive when compared to the response times of photoconductive or bolometer-based devices. The silicon photodiode has a responsivity ca 0.4 A/W at the peak response wavelength. Some manufacturers list NEP values of less than 2×10^{-15} W Hz^{1/2} and D^* above 1×10^{14} cmHz^{1/2}/W and optical areas of 1–10 mm². The photodiode has proven to be a useful tool for photon-counting and imaging applications over the entire range of spectral sensitivity and has been utilized in the visible portion of the spectrum for power generation (56,57).

Silicon photodiodes are available in both discrete and array formats. The simplicity of the discrete photodiode makes this device one of the least expensive photon detectors available. Discrete photodiodes are merely p - n homojunctions and are available from numerous manufacturers. Typical commercially available devices have one to four discrete diode detectors per package and frequently device manufacturers include operational amplifier-based readout electronics inside the packaged device for ease of use. Discrete photodiodes can be used in a myriad of applications including high speed optical switching, intensity determination for automatic exposure control circuitry in film cameras, and photon counting for spectroscopic analyses (58).

Photodiode arrays are more complex than their discrete counterparts due to the difficulty of directing the signal information from each diode to off-chip electronics. The more common linear arrays contain internal multiplexing circuitry located on the periphery of the imaging area. This circuitry amplifies and buffers the signal from each diode, presenting the information from each pixel through a single output amplifier in a controlled, time-sequenced fashion. The performance characteristics of linear photodiode arrays typically rival those obtained from discrete diodes. Arrays ranging in size from 1×64 pixels up to 1×4096 pixels and larger are available from commercial sources. Linear photodiode arrays are commonly found in high resolution image scanning applications such as photocopiers and facsimile (FAX) machines.

As for linear photodiode arrays, two-dimensional photodiode arrays require internally integrated circuitry to mediate the signal information from each pixel. However, with the exception of the outermost rows and columns, the pixels in two-dimensional arrays are surrounded on all sides by other pixels. Thus the required circuitry cannot reside solely on the periphery of the array but must be integrated into the actual pixel site. In a two-dimensional array each pixel consists of both a photodiode and electronic circuitry designed to attach or detach that diode from the readout electronics located at the chip periphery. The voltage pulses which control the time-sequenced readout are generated from two multiplexing circuits, one each for the x - and y -chip dimensions, which are also located at the chip periphery. In its simplest form, the two-dimensional photodiode array utilizes a large capacitor that is common to all pixels in a given row to sequentially convert the signal charge from each pixel to a voltage. The resulting signal voltage is very small, resulting in a signal-to-noise ratio roughly one-half that of an equivalent format CCD array. Generally device performance varies inversely with the number of pixels in the array. Prior to ca 1984, two-dimensional photodiode arrays were heavily utilized in commercial imaging applications such as video cameras. Recently CCDs and metal oxide semiconductor (MOS) arrays, two-dimensional imaging arrays which utilize p - n diodes as photosites, and complimentary metal oxide semiconductor (CMOS)-based components for readout circuitry, have emerged as strong competitors in this arena. Nevertheless some two-dimensional standard video format photodiode arrays are still manufactured. These devices are most useful in situations unsuited for CCD and CMOS-based imagers, such as ionizing radiation environments.

In addition to being a popular image detector, the silicon homojunction photodiode and avalanche photodiode (59) can also be used for power generation (qv) by operating the device in the photovoltaic mode (see PHOTOVOLTAIC CELLS). In this mode incident photons produce a voltage drop across the device that is proportional to the number of absorbed photons. When a finite load resistance is placed across the diode leads, a current is produced. Thus, silicon homojunction diodes can be used to convert optical energy into electrical energy. The energy conversion efficiency for common silicon photocells is in the 3 to 15% range depending on the specifics of the device design. High levels of illumination are required to produce useful output power. For typical photosensitive cells an illumination of 10 lux can produce an open-circuit output voltage of ca 0.5 volts. However, any desired voltage or current can be generated by the appropriate series or parallel interconnection of multiple individual elements. Whereas silicon photodiodes are not as efficient at power generation as more exotic Group 2–16 (II–VI) and Group 13–15 (III–V) materials, these devices are commonly utilized as power sources for both terrestrial and space-borne applications.

Fabrication. Photodiodes are made by a process similar to that used to manufacture bipolar integrated circuits (qv). For a p - on n -diode formation process, the starting wafer is an n -type silicon substrate of roughly 500- μ m thickness. The silicon has been doped with either P, As, or Sb during the wafer formation process to a nominal resistivity of 10 Ω cm. The initial step of the fabrication process consists of growing a thick layer of SiO₂ on the top surface of the silicon wafer through the use of elevated temperatures in a steam and oxygen ambient. Next, circular windows are etched in the SiO₂ layer. The wafer is then placed in a high temperature diffusion furnace which introduces boron dopant into the silicon through the open circular windows. Ion implantation (qv) is a common alternative to high temperature diffusion as a means of boron doping. The result of the doping procedure is the formation of a p - n junction, with one diode formed for every window in the SiO₂. Next a low resistance material such as aluminum is used to form ohmic contacts to the p -type silicon regions. The aluminum layer is also used to form distinct electrodes, one per diode, to which external connections can be made. Finally a single common ohmic contact is formed on the back surface of the n -type silicon wafer by an additional metal layer.

The method of diode formation as well as the density and profile of the impurity ions determines the specific optical and electrical performance parameters of the photodiode. When photon absorption occurs in the depletion region of the diode, the resulting carriers are quickly swept from the diode and measured by the readout circuitry. The same behavior occurs for optically induced charge formed within a diffusion length of the depletion region. Charge generated at a distance greater than one diffusion length recombines in the undepleted silicon and consequently cannot be detected as signal charge. A similar fate befalls photogenerated carriers produced in heavily doped or heavily lattice-damaged regions. Heavy surface doping and lattice damage are common by-products of the homojunction formation process. Therefore, the diode fabrication process balances the reduction in the rate of surface dark current generation against the charge collection loss produced by heavily doping the silicon surface. Similarly, the improvement in photon collection efficiency obtained by increasing the thickness of the depletion region must be weighed against the associated increase in bulk-generated dark current.

Cadmium Sulfide Photoconductor. CdS photoconductive films are prepared by both evaporation of bulk CdS and settling of fine CdS powder from aqueous or organic suspension followed by sintering (60,61). The evaporated CdS is deposited to a thickness from 100 to 600 nm on ceramic substrates. The evaporated films are polycrystalline and are heated to 250°C in oxygen at low pressure to increase photosensitivity. Copper or silver may be diffused into the films to lower the resistivity and reduce contact rectification and noise. The copper acceptor energy level is within 0.1 eV of the valence band edge. Sulfide vacancies produce donor levels and cadmium vacancies produce deep acceptor levels.

The settling technique can be accomplished from an ink which contains 1- μ m crystallites of CdS and selected concentrations of CdCl₂. The coating is fired in a restricted volume of air at 500–700°C. During the sintering, the CdCl₂ acts as a flux and forms a solution with the surface of the grains of the CdS. A few ppm of chloride and copper (from impurities) enter the crystal lattice and act as activator centers for trapping. Excess chloride evaporates leaving a continuous polycrystalline photoconductive film. The layers are from 5 to 30- μ m thick and have a linear I - V relationship.

The films have an area resistance of 100,000–300,000 Ω square. Most applications, such as switching on outdoor lights at twilight, require a detector resistance of near 1000 ohms to operate the switching circuit without the need of impedance matching electronics. This is accomplished by depositing the contacts in an interdigitated geometry as shown in Figure 12. A protective film is deposited over the detector and contacts to provide for long-term stability or the detector structure is mounted in a hermetic package as shown. The spectral sensitivity for the thin-film CdS detector is shown in Figure 9a. The response shape is similar to that of the human eye (see Table 1).

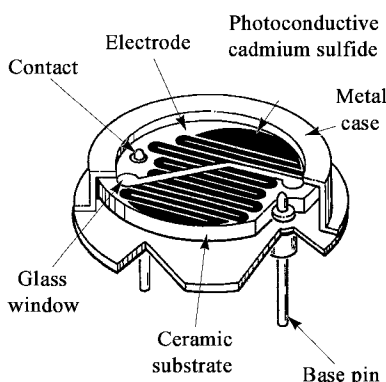


Fig. 12. CdS film detector in a package showing interdigitation to reduce resistance. See text.

GaAsP and InGaAs Photodiodes. Gallium–arsenic and gallium–arsenic–phosphorus diodes are fabricated as photodiodes as well as light emitters (see **LIGHT GENERATION, LIGHT EMITTING DIODES**). Fabrication is typically with mesa etch technology of the films of GaAsP or InGaAs grown (62) by the vapor-phase epitaxial process using metal organic chemical vapor deposition (MOCVD) (see **THIN FILMS, FILM FORMATION TECHNIQUES**). This growth technique results in impurity densities less than 1×10^{14} atoms cm^{-3} . The spectral cutoff range extends from 500 to 900 nm depending on the phosphorus (qv) content. These detectors can be used for color discrimination and do not require expensive interference filters. Emitter–diode pairs are utilized for very high impedance signal coupling in high speed integrated circuits. Spectral sensitivity is shown in Figure 9a for typical GaAs and GaAsP photodiodes. Surface leakage is probably a limitation because of the lack of a native surface-passivation technology such as SiO_2 for Si detectors. The

indium–gallium–arsenide, InGaAs, photodiode has made a comeback with improved epitaxial growth techniques. The available cutoff wavelength is 1.7 μm and is advertised as a linear array for spectrometers and environmental monitoring (see Table 1).

PbS and PbSe Photoconductors. The lead chalcogenides, PbS, PbSe, and PbTe, were among the first infrared-detector materials to have been investigated. Although photovoltaic effects are observed with p – n junctions in single-crystal material the response is quite poor and not reproducible. However, very sensitive photoconductors are prepared as polycrystalline thin films, ca 1- μm thick, which are deposited on glass (qv) or quartz substrates between gold or graphite electrodes.

Detector elements are prepared either by sublimation in the presence of a small partial pressure of O_2 or by chemical deposition from alkaline solution containing a lead salt and thiourea or selenourea (63). Lead sulfide and lead selenide deposit from solutions as mirror-like coatings made up of cubic crystallites 0.2–1 μm on a side. The reaction may nominally be represented by the following:



The actual reaction probably is more complex. The photoconductive behavior depends on the pH of the solution from which deposition occurs. It is likely that oxygen-containing compounds are present in the deposited films. For either method of preparation, the effect of oxygen, which is introduced during preparation or by subsequent heat treatment in air or oxygen, is critical for the development of optimum sensitivity. Maximum sensitivity is obtained near the point at which the film conductivity type changes from n to p . The long response times are suggestive of trapping. It is likely that deep trapping states are located at the oxidized surface of the micrograins (64). The evaporation technique produces the best results especially for PbSe and the more obsolete PbTe. The spectral sensitivity is shown in Figure 9b for different operating temperatures (see Table 1).

Platinum Silicide Schottky Barrier Arrays. The Pt:Si detector (65–67) essentially is a metal semiconductor barrier whereby the platinum silicide is a quasi-metal that generates a small energy barrier to electrons. The effective photons are absorbed in a very thin region of the silicide next to the barrier and generate free electrons that flow over the barrier and tunnel through it into the n -type silicon. The efficiency of this process is only a few percent even at high energies (short wavelengths) because of the low electron diffusion coefficient in the silicide. At wavelengths beyond 1 μm the efficiency drops off dramatically because of decreasing tunneling probability for the lower energy electrons. The effective quantum efficiency is ca 0.1% at 4.8 μm wavelength.

Techniques of platinum deposition vary but sputtering and annealing of a very thin layer of platinum produces a uniform platinum silicide film with less than 0.3% variation in responsivity in an area of 2 $\times$ 2 cm. Details of deposition and annealing processes are considered trade secrets by the manufacturers. Deposition is typically onto silicon integrated circuit (IC) chips having a readout structure (see **INTEGRATED CIRCUITS**). Small regions of the IC chip at each pixel are dedicated for the silicide detector element. The readout is typically an in-line charge-coupled device. The spectral sensitivity is shown in Figure 9b and array information is given in Table 1.

Although the photon efficiency is quite low, focal plane performance for infrared imaging of ambient temperature scenes is acceptable because of long television display frame time, reasonably low detector noise, and the excellent uniformity of responsivity which allows for high on-chip input gain without offset correction. The responsivity is ca 10 millivolts per degree delta scene temperature using $f/2$ optics and scene sensitivity with large arrays is ca 0.1°C.

InSb Photodiode Detectors and Arrays. Sensitive photodiodes (68,69) have been fabricated from single-crystal InSb using cadmium or zinc to form a p -type region in bulk n -type material. High quality InSb crystals can be grown by the infinite-melt process (70) where an InSb film is grown epitaxially (from the liquid phase) on a slice of InSb which was prepared in a conventional Czochralski vertical puller. The diode formation process typically is a closed-tube diffusion. Cleaned and etched samples of InSb are placed in a quartz ampul with a limited amount of zinc or cadmium. After evacuation and sealing, the ampul is heated to ca 50°C below the crystal melting point. The metal vaporizes partially or completely, depending on the amount, volume of ampul, and temperature, and diffuses into the crystal after a few hours. The impurity–diffusion profile approximates the error–function law and the p – n junction is 1–5 μm below the surface. Diode arrays (16) are formed by etching mesas ca 50 μm square. Typical performance of InSb detectors is given in Table 1 and the spectral sensitivity is shown in Figure 9c. Up to 480 $\times$ 640 matrix arrays of InSb photodiodes in a mesa configuration have been demonstrated. Commercial units of 256 $\times$ 256 are available. The mesa detector array is mated to a silicon chip having an array of amplifiers and multiplex circuitry. Each diode is connected to an amplifier input. The hybridization process consists of forming indium bumps on each diode mesa and on each amplifier input, using a photolithographic process and In evaporation and pressing the detector array chip to the silicon integrated circuit chip. The infrared must pass through the InSb chip to reach the photodiode junction. To improve quantum efficiency the InSb is grown on GaAs or GaAsSb substrates as a thin layer. Quantum efficiency is greater than 50% for wavelengths greater than 2 μm and less than the cutoff of InSb, 5.3 μm . A protective coating (passivation) of the InSb photodiode for stable operation over several hours without frequent signal normalizations has not been found. However, infrared imaging cameras using hybrid InSb focal planes are a commercial reality. In a real-time imaging configuration the scene sensitivity is ca 0.04°C using an InSb infrared camera.

Mercury Cadmium Telluride. HgCdTe has proven to be an excellent infrared detector material (2) where the CdTe content can be readily adjusted to obtain cutoff wavelengths from 2 to 20 μm . The benefit is high spectral sensitivity of the photon detector, low defect density, and high cooling efficiency. The dependence of energy gap on mole fraction is linear. For $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$, the x values of most interest lie between 0.17 and 0.50. The need for large focal planes up to 2 $\times$ 2 cm has dramatically changed the direction of single-crystal HgCdTe growth technology.

Crystal Growth. The method of solid-state recrystallization (quench and anneal) was marginally adequate during the 1970s and 1980s for linear photoconductor arrays used in military night vision systems, but the quenching kinetics restricted sample size to 6 mm $\times$ 2 cm. The crystals had many low angle grain boundaries, defect densities were high, and there were large nonuniformities in composition and carrier concentration. Therefore epitaxial

growth technologies were developed. Crystal growers were hampered, however, by the lack of suitable substrate material until methods were developed to grow the HgCdTe films onto large-area silicon and GaAs single crystals. The problem with misfit dislocations has not as of this writing (ca 1995) been satisfactorily solved except for detector cutoff wavelengths less than 5 μm . Thus, CdTe single crystals are used to obtain large, low defect density substrates for growth from the liquid phase. Lattice matching is achieved by adding 3% Zn by atomic weight.

Large-area CdZnTe substrates form the basis for liquid-phase epitaxy (LPE) growth of mid- and long wavelength ir HgCdTe detector material. Substrates are obtained from 3.5 kg CdZnTe ingots grown in graphite boats in sealed quartz ampuls (71). The horizontal Bridgman growth process yields $41 \times 6.4 \times 5.0$ cm semicylindrical ingots having large single-crystal portions which are then sectioned, sawed into slabs, and diced into required dimensions. The substrate surfaces are diamond turned and etched to assure flat, damage-free surfaces. Standard substrate sizes are 3.6×2.0 cm and 3.6×1.5 cm, but the ability to obtain single-crystal $\langle 111 \rangle$ oriented CdTe substrates, as large as 5.1×7.6 cm, has already been demonstrated. The material typically exhibits dislocation densities of 1×10^5 cm^{-2} and high purity as judged by Hall measurement made on evaluation samples of *n*-type LPE films grown on these substrates.

Liquid-phase epitaxial films (71,72) are grown in production prototype dipping reactors as shown in Figure 13 using the CdZnTe substrates. Film growth is both from tellurium and mercury solutions. Phase diagrams for the Te and Hg corners are shown in Figure 14. Growth is from lightly (ca 5×10^{14} cm^{-3}) indium-doped 4000-g tellurium solutions. The mercury vapor pressure is maintained by a mercury reservoir positioned in an independently controlled furnace zone. Up to 54 cm^2 of material can be grown from the largest reactors in a single run. Multiple furnace zones in the vicinity of the melt crucible permit adjustment of the melt temperature profile. The substrates are positioned horizontally during growth and rotated to promote uniformity of composition. The holder design allows substrates to be reoriented vertically for withdrawal, to facilitate melt drainage. LPE films are also grown in mercury solutions of several kilograms in which small amounts of tellurium and cadmium have been added. In both cases the cutoff wavelength varies less than 0.1 μm across the entire film.

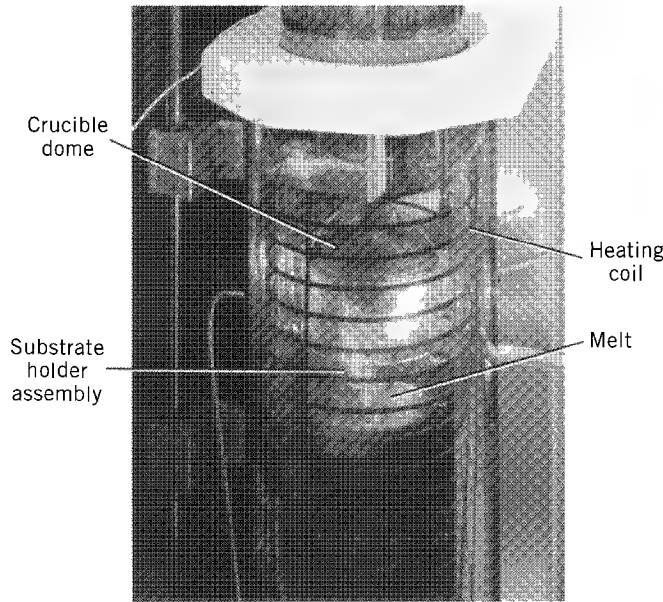


Fig. 13. Furnace reactor for the growth of HgCdTe films on CdZnTe substrates using the liquid-phase epitaxial process. The melt is tellurium in a quartz crucible. A mercury reservoir in a cooler zone maintains the required Hg over-pressure.

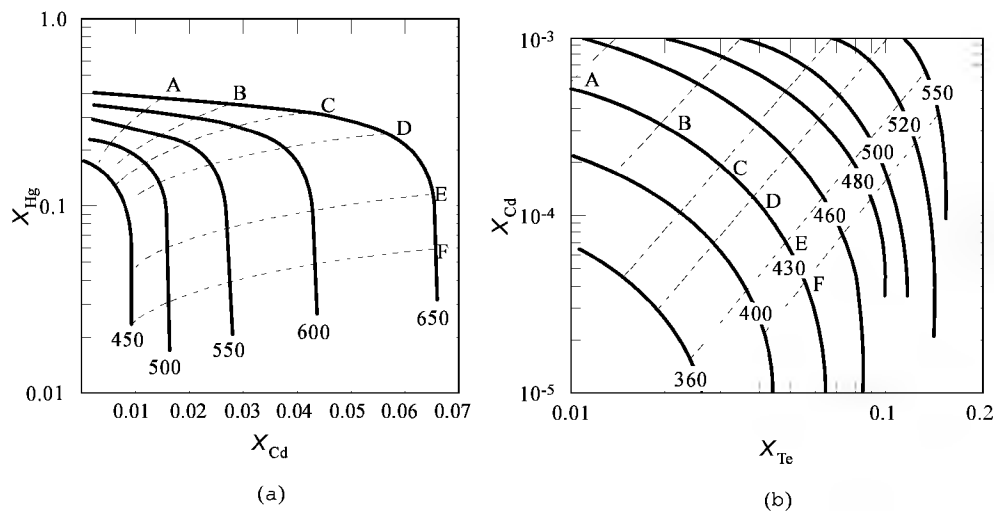


Fig. 14. Phase diagrams of HgCdTe used to define the liquid-phase epitaxial growth process where composition is in mole fraction, *X*, and the numbers represent temperatures in °C: (a) Te-rich corner where the dotted lines A–F correspond to values of *X*_{Te} of 0.1, 0.2, 0.3, 0.5, 0.8, and 0.9, respectively, and (b) Hg-rich corner where A–F correspond to values of *X*_{Hg} of 0.9, 0.8, 0.6, 0.4, 0.2, and 0.1, respectively.

The films grown in tellurium are annealed in Hg vapor. Mercury atoms diffuse into the films to reduce the density of Hg vacancies which act as acceptor sites. The carrier concentration is shown as a function of anneal temperature in Figure 15. For photoconductive detectors the films are annealed such that the indium donor density is dominant resulting in *n*-type material having a 1×10^{15} cm^{-3} excess electron density. Minority carrier lifetime is typically 1–5 μs at 77 K. For photodiodes the films are annealed to an acceptor (via Hg vacancies) density of 3×10^{16} cm^{-3} . Carrier lifetimes are less than 50 ns at 77 K. Extrinsic acceptor doping is being developed to replace vacancy doping to achieve longer minority carrier lifetime and lower dark current density. Film thickness is 40–80 μm .

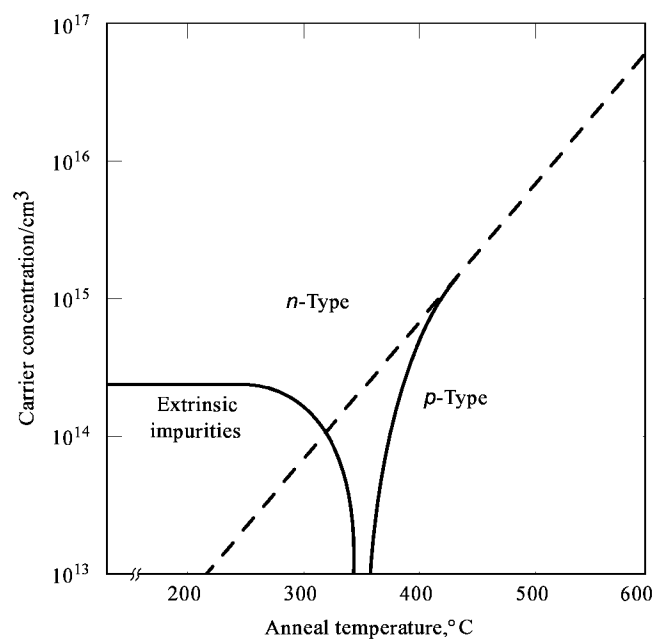


Fig. 15. Excess carrier concentration in HgCdTe in a saturated Hg vapor as a function of temperature where the dashed line represents Hg vacancies. The extrinsic impurity concentration can be adjusted in the growth process from low 10¹⁴ up to mid-10¹⁷. Low temperature annealing reduces Hg vacancy concentration and acceptor density.

Films grown in Hg (72,73) are usually structured to make heterojunction photodiode arrays. The first or base layer is narrower band gap HgCdTe, grown on CdZnTe substrates, doped with indium for excess electrons (*n*-type) in the 3 × 10¹⁶ cm⁻³ range and is 10-μm thick. The second or cap layer is wider band gap HgCdTe, doped with arsenic for excess holes (*p*-type) in the 5 × 10¹⁵ cm⁻³ range and is 4-μm thick. The composite HgCdTe film is photolithographically etched part way into the base layer to form an array of mesas, each one being a photodiode detector element. The *p-n* junction is close to or coincident with the metallurgical heterojunction. For infrared detection in the 8–12 μm atmospheric spectral window the base layer CdTe content is ca 20% and the cap layer CdTe content is ca 30%.

Detectors Arrays. Greater than 50,000 HgCdTe linear arrays (74,75) have been produced in the United States since 1972 for the Department of Defense infrared systems ranging from night vision for M1 tanks to targeting sights for laser guided bombs. These common module detector arrays consist of 180 elements photoetched on a 50-μm pitch. Virtually all of the material used for these arrays was prepared by the solid-state recrystallization process. Small slabs of material were epoxied to sapphire or ceramic and were thinned to 8 μm. The newer epitaxial HgCdTe is also epoxied to the ceramic with the film side down. The material is thinned by diamond turning and the CdZnTe substrate is removed. In each case an 8-μm thick layer of HgCdTe remains. The last few micrometers of removed HgCdTe must be carefully etched to avoid generating defects. The arrays are defined by photoetching, passivated using a ZnS film, and indium electrical contacts are applied. The contacts define the optically active area of ca 2 × 10⁻⁵ cm². The arrays are mounted in a vacuum Dewar for cryogenic cooling. Each element is connected to a series resistance bias circuit and a-c coupled to an amplifier. Detector resistance is nominally 100 ohms and bias current is ca 2 mA. The signal voltage is given by equation 21 and the responsivity by equation 34.

$$R_V = \eta \tau V \ h \nu n v$$

(34)

The responsivity becomes independent of the bias voltage, *V*, when the electric field-induced sweep time of the holes equals the hole lifetime.

For a well-designed, well-made HgCdTe photoconductor detector (76,77), g-r noise is dominant and may be expressed in terms of a minority carrier density *p* and majority carrier density *n*. Semiconductor noise analysis for the HgCdTe photoconductor yields,

$$V_{g-r}^2 = \frac{4p\tau V^2 \Delta f}{n(n+p)v}$$

(35)

The responsivity and g-r noise may be analyzed to obtain background photon flux and temperature dependence of responsivity, noise, and detectivity. Typically, *n* > *p*, and both are determined by shallow impurity levels. The minority carrier density is the sum of thermal and optical contributions,

$$p = \frac{n_i^2}{n} + \frac{\eta \phi_B \tau}{t}$$

(36)

where, typically, the background flux ϕ_B is much greater than the signal flux ϕ_s . The lifetime, τ , is ca 1 μs but for narrow band gap defect-free detectors it becomes the Auger lifetime and can be calculated readily from basic semiconductor properties (74). The cooling requirement is determined according to equation 36 because *n_i* is an exponential function of band gap energy and temperature. Combining equations 35 and 36 detectivity can be expressed as follows:

$$D_\lambda^* = \frac{\eta}{2h\nu} \left(\frac{\tau}{pt} \right)^{1/2}$$

(37)

At low background flux this gives the temperature dependence of the *D** shown in Figure 4. At high flux, the *D** equation (eq. 37) reduces to equation 12 except for a factor of (2)^{1/2} which is a result of the random recombination process not present in diodes. The scene sensitivity of a scanning photoconductor array infrared camera is ca 0.15°C.

The long minority carrier lifetime in *n*-type HgCdTe has been exploited to perform signal processing in the element (SPRITE) (78). The bias field is adjusted to sweep the photogenerated holes along the HgCdTe detector element synchronously with the scanned image creating a time delay and integration (TDI) enhancement of the signal-to-noise ratio.

Detectors, Arrays, and Focal Planes. The two popular types of photodiode arrays in HgCdTe are based on homojunction (79) and heterojunction (80,81) technologies. Homojunction diode arrays are fabricated with *p*-type epitaxial HgCdTe. The surface of the grown HgCdTe film is flattened by diamond milling and chemical lapping using a weak solution of bromine in methanol. The epitaxial film is then epoxied to a silicon chip having an array of amplifiers (typically on 50-μm centers) and readout multiplexer (mux) circuitry. The CdZnTe substrate is milled away and the HgCdTe is thinned to ca 10 μm. An array of small (10–20 μm) holes is etched in the HgCdTe film such that each hole is located over an amplifier input pad of the

silicon IC chip. The diodes are made by implanting boron ($150\text{ KV}, 1 \times 14\text{ cm}^2$) through a 500-nm ZnS layer. Planar diodes are generated using patterned photoresist to define the implanted regions located between the holes. The implant process disrupts the lattice, creating Hg ion interstitials which cause the formation of very shallow donor states. The damage layer extends ca 150 nm from the surface and the $n\text{-}p$ junction varies from 2 to $8\text{-}\mu\text{m}$ deep depending on the implant and annealing conditions. The fabrication is completed by applying a CdTe passivation layer (ca $1\text{-}\mu\text{m}$ thick) directly to the HgCdTe surface, forming a ZnS layer for antireflection and electrical isolation and formation of a metal film lead to connect each implanted n region to each amplifier input. Infrared illumination is directed onto the HgCdTe film and quantum efficiency typically exceeds 75% . Because infrared focal planes are typically cooled to below 100 K the differential coefficient of thermal expansion causes the shrinking silicon to put tensile stress on the HgCdTe film. However, the thinness of the film and the presence of the holes allow the HgCdTe to strain retaining the integrity of the epoxy film and the electrical contacts. Array dimensions up to 2.5 cm have proven reliable.

Heterojunction diode arrays utilize the grown $p\text{-}n$ junction and mesa etch technology. The mesa arrays are passivated by a thin layer (ca 500 nm) of CdTe using an evaporation or molecular beam epitaxial process. Annealing at 250°C in a Hg atmosphere creates a short (ca 100 nm) graded layer from the CdTe to the HgCdTe. The benefit of CdTe passivation is efficient isolation of the $p\text{-}n$ junction from the surface, low dark current noise, and immunity to ionization radiation. As for the InSb hybrid arrays, indium bumps are formed using evaporation at the center of each mesa and on each amplifier contact pad of the silicon integrated circuit chip. The two chips are pressed together in a bump bonding process called hybridization (82). Sometimes the space between the chips is filled with epoxy. Infrared illumination is through the CdZnTe substrate which has been coated with ZnS for antireflection. Quantum efficiency typically exceeds 75% .

HgCdTe photodiode performance for the most part depends on high quantum efficiency and low dark current density (83,84) as expressed by equations 23 and 25. Typical values of R_0A at 77 K are shown as a function of cutoff wavelength in Figure 16 (70). HgCdTe diodes sensitive out to a wavelength of $10.5\text{ }\mu\text{m}$ have shown ideal diffusion current limitation down to 50 K . Values of R_0A have exceeded $1 \times 10^6\text{ }\Omega\text{cm}^2$. Spectral sensitivities for three compositions of HgCdTe detectors are shown in Figures 9a and 9c. More information is listed in Table 1. The scene sensitivity with a HgCdTe diode area array cooled to 80 K is ca 0.02°C for ambient temperature scenes.

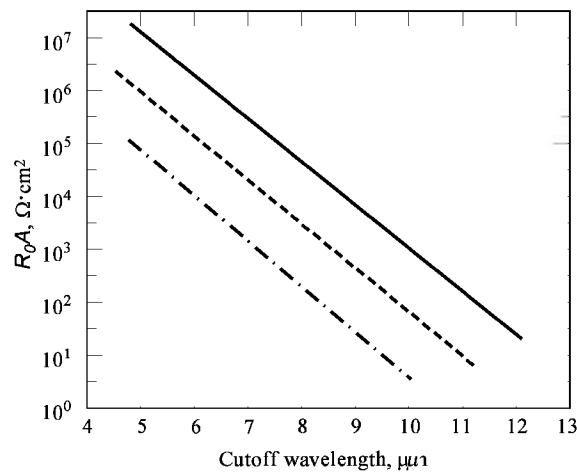


Fig. 16. Resistance area (R_0A) product for HgCdTe photodiodes cooled to 77 K . The solid line represents the theoretical limit, the dashed lines () and (•) high and low performance, respectively. Dark current caused by defects lowers R_0A and detector sensitivity. In the high performance range dark current is an exponential function of cutoff wavelength and temperature.

Doped Germanium and Silicon Photoconductors. The extrinsic photoconductors are typically single-crystal germanium doped with zinc, cadmium, mercury, boron, and gold (85,86) and silicon doped with indium, gallium, or arsenic (87). The doping density ranges from 1×10^{15} to 1×10^{17} impurity atoms cm^{-3} leading respectively from low (1 cm^{-1}) to higher absorption coefficient (50 cm^{-1}).

Information on ionization energies, solubilities, diffusion coefficients, and solid-liquid distribution coefficients is available for many impurities from nearly all columns of the Periodic Table (86). Extrinsic Ge and Si have been used almost exclusively for infrared detector applications. Of the impurities, Cu, Au, Zn, Cd, Hg, and some of the elements of Groups 13 (III) and 15 (V) have been used in detectors. Germanium and silicon, which are used in the preparation of detectors, must be of high purity before they are doped with the desired activator impurity in order to avoid unwanted compensation by impurities with smaller ionization energies than the activator. The required purity can be achieved by zone refining in which a short molten zone is repeatedly passed from end to end of an ingot of impure Ge or Si. Impurities having distribution coefficients larger than unity collect near the seed. The concentration of electrically active residual impurities in the center portion of the ingot can be reduced to $10^{12}\text{--}10^{13}\text{ cm}^{-3}$. Single crystals can be grown using the Czochralski method, in which an oriented seed crystal is brought into contact with the melt and then is withdrawn slowly while being rotated, or using a horizontal zone melting method, in which a seed crystal is melted onto one end of a polycrystalline ingot. A molten zone is produced at the junction of the ingot and seed and is moved slowly along the ingot, leaving behind a single crystal. All of these operations must be carried out in an inert or reducing atmosphere in order to prevent oxidation of the germanium or silicon.

In most cases, the activator impurity must be incorporated during crystal growth. An appropriate amount of impurity element is dissolved in the molten Ge and, as crystal growth proceeds, enters the crystal at a concentration that depends on the magnitude of the distribution coefficient. For volatile impurities, eg, Zn, Cd, and Hg, special precautions must be taken to maintain a constant impurity concentration in the melt. Growth occurs either in a sealed tube to prevent escape of the impurity vapor or in a flow system in which loss caused by vaporization from the melt is replenished from an upstream reservoir.

Some impurities, eg, Cu, Ag, Ni, Co, and Fe in Ge, and In, Ga, and As in Si, have diffusion coefficients that are large enough to permit doping by solid-state diffusion at well below the melting point of the host crystal. A thin layer of the diffusant, which is deposited on the surface of the crystal by electroplating, vacuum evaporation, or electrochemical replacement, serves as the source for diffusion. After homogeneity is achieved, the sample is quenched. The alloyed surface layer must be removed by lapping and etching before the electrical contacts are applied. The impurity concentration should be as large as possible, within limits, in order to maximize the absorption coefficient. In some cases, the concentration is limited by the impurity solubility, which may be small depending on the element. For impurities of high solubilities, the upper limit is set by the onset of impurity banding. When the average separation of impurity atoms in the lattice is small enough, conduction by direct transfer of carriers from atom to atom can occur. Impurity banding limits the extent by which cooling can reduce the dark current and, therefore, the noise, in Ge and Si. This is significant above impurity concentrations of ca 10^{17} cm^{-3} .

Impurities other than those from Groups 13 (III) and 15 (V) generally exhibit two or more impurity levels in Ge. If an activator which is not of the lowest lying level is used, the lower lying levels must be compensated for by addition of an impurity of the opposite conductivity type. For example, if the second Zn acceptor level for 0.095 eV is to be used, the lower lying level at 0.035 eV must be compensated for by addition of a donor impurity, eg, As, in a concentration slightly greater than that of Zn. Electrons from the As donors fall into the low lying Zn level and render it inactive. A special case is the lowest lying Au level; Au acts as a donor at 0.045 eV above the valence band. If a shallow acceptor, eg, Ga, is added in a concentration that is slightly less than that of Au, electrons from the donor level fall into the Ga level. At low temperatures, holes which are bound to the compensated Au centers can be

photoexcited to yield photoconductivity with a long wavelength threshold at ca 25 μm . Single-detector elements and arrays are formed by dicing and etching and attachment of electrical contacts. Linear arrays of several hundred elements have been made. The spectral sensitivities of doped Si and Ge detectors are shown in Figures 9c and 9d. See Table 1 for other information. Monolithic and hybrid extrinsic focal planes have also been demonstrated (88,89).

GaAs–AlGaAs Quantum Well Arrays. The quantum well infrared detector (90,91) is a newer technology based on the artificial structure called a superlattice. The quantum well detector array is an attempt to achieve a monolithic architecture and thereby very large size with high reliability. The idea is to build the long wavelength detector array directly on the readout integrated circuit. Infrared detection out to 12- μm wavelength can be achieved using engineered material having a controlled energy gap. The technique is to fabricate multilayers of semiconductors having alternating band gaps such that a series of potential wells exists in the direction normal to the layers. Some success has been achieved using very thin layers of GaAs and AlGaAs grown by molecular beam epitaxy (MBE). The MBE process is conducted in stainless tubing and is simply controlled evaporation of the elements from side chambers into a main chamber containing a wafer of the substrate material, typically GaAs. The electron wave function is confined in the potential wells when the photon-generated electron wavelength is nominally the layer thickness (ca 10 nm). The electrons are freed from the wells by photons and a signal is detected with the application of an external bias voltage.

AlGaAs quantum well infrared photodetector (QWIP) focal planes have achieved sufficient sensitivity out to 10- μm wavelength to result in scene temperature sensitivity of ca 0.2°C when the focal plane is cooled to 77 K. Spectral sensitivity is shown in Figure 9c and array information is given in Table 1. The superlattice, a newer tool for achieving controlled activation energy, should present many alternative infrared detection techniques.

Semiconductor Bolometer Arrays. The use of bolometers to sense infrared radiation is not new. What is different is that, rather than having a single cell of large dimension, a matrix of miniature cells or microbolometers is created. Rather than a single amplifier connected externally, a custom integrated circuit is built under each cell to form a totally integrated focal plane array as indicated in Figure 17. Key elements of the structure (92,93) are the thin (100 nm) amorphous silicon ($\alpha\text{-Si}$) or vanadium oxide (VO) thermally sensitive membrane, the thermally insulating support arms, and the integrated circuit underlying this structure. Infrared energy focused on the individual pixels heats the membrane as given by equation 24. The support arms act as electrical contacts but are sufficiently thin and narrow to prevent significant conduction of thermal energy from the membrane to the surroundings. Temperature-dependent electrical resistance of the semiconductor film is monitored by the circuitry and converted into an electrical signal proportional to the incident radiance. This signal is displayed on a standard television monitor, thus forming the image of the scene.

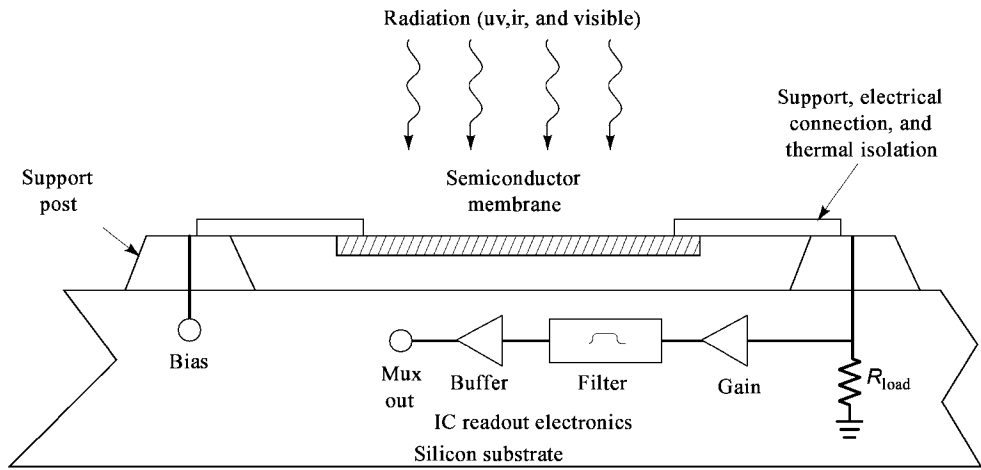


Fig. 17. Cross-sectional schematic of a microbolometer photodetector. Micromachining is used to construct small, very low mass detectors. The dimension of a detector–amplifier cell is 50 μm . Mux = multiplexer, IC = integrated circuit.

Two technological advances have occurred that make such a structure feasible to build. The first is the development of microetching techniques that can be used to form microscopic structures in silicon and its coatings. This makes formation of the membrane and support arms possible with thickness control in the 10-nm range. Cell dimensions of less than 50 μm have been demonstrated. Scanning electron microscope photos are shown in Figure 18 of an array and a test pixel. The second critical factor is the increased circuit density in silicon integrated circuits, making it possible to build a small circuit for each cell of the array. By doing this, the noise bandwidth of each cell can be minimized, thereby maximizing performance. This device is a thermal detector in the infrared and is therefore independent of wavelength. Measured responsivity is ca $5 \times 10^5 \text{ V/W}$ and NEP is ca 50 pW. The focal plane operates at ambient temperature. Imaging arrays having scene sensitivity better than 0.1°C have been demonstrated (see Table 1).

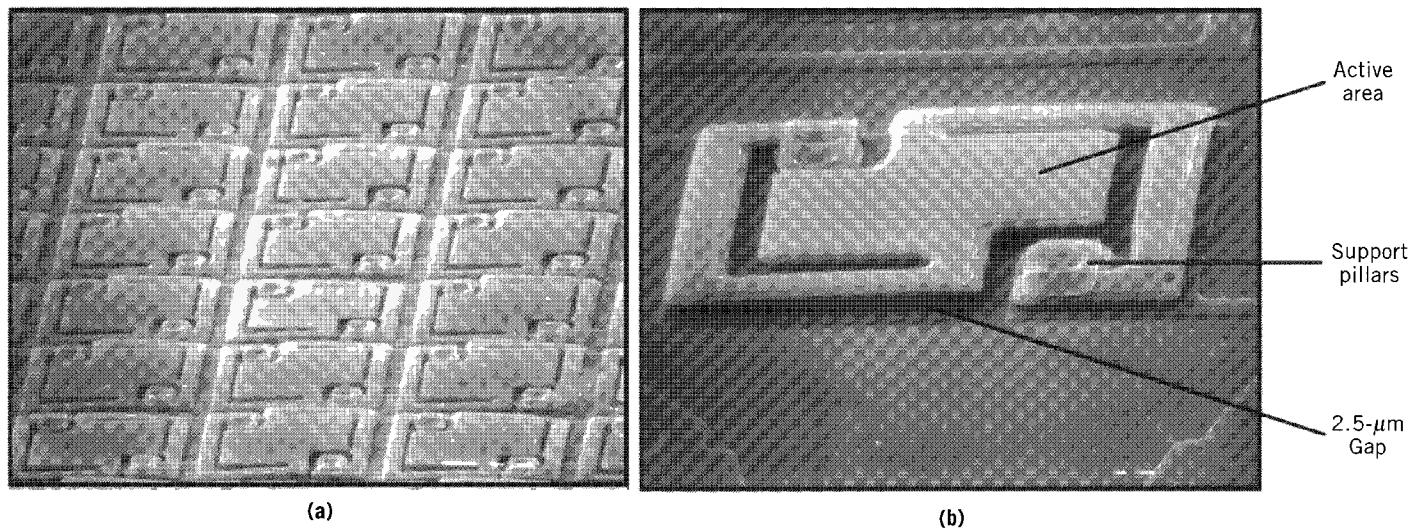


Fig. 18. Microbolometer (a) array portion showing pixels on a 50- μm pitch. Each pixel is connected to a readout amplifier in the supporting silicon IC chip. (b) Detector having a 35 $\times$ 40 μm active area. The serpentine arms give excellent thermal isolation and the low mass results in a 10-ms response time, ideal for thermal imaging.

Economic Aspects

The cost of a photodetector depends on such factors as material preparation yield, complexity of fabrication, packaging, and demand. Detector prices in the United States for 1980 and 1994 are listed in Table 2. The fairly constant prices of silicon devices have resulted primarily from increased automation. Devices which can be cooled are very expensive because they operate close to the theoretical limit, and packaging for long-term, reliable operation is difficult (see PACKAGING, ELECTRONIC MATERIALS). Integrated focal plane arrays are complex, having entered the open market in 1993. CCD arrays for video cameras, a high production item, are made in sophisticated silicon front ends where yields are high. The U.S. government is funding efforts to lower the cost of infrared focal planes by the late 1990s for commercial uses.

Table 2. Photodetector Prices in the United States, \$^a

Detector	1980	1994
<i>Uncooled</i>		
Si diode element on header	3	1
Si CCD for video camera	250	50
CdS photoconductor	1	1
GaAsP diode on header	15	20
PbS detector on header	2	2
microbolometer focal plane (240 × 320)		5000
<i>Cooled in Dewar flask</i>		
InSb diode	1100	500
InSb focal plane (256 × 256)		5000
Pt:Si focal plane (480 × 640)		500
HgCdTe detector element	2000	800
HgCdTe photoconductor array (1 × 180)		1200
HgCdTe diode focal plane (128 × 128 and 480 × 4)		6000

^a Prices depend strongly on specifications of sensitivity and quantity.

Health and Safety Factors

The completed photodetector usually is packaged hermetically in inert glasses or plastics or is enclosed in an evacuated metal or glass container. Although most detector materials are toxic, the means taken to passivate and isolate these materials are often adequate to protect the user. However there are exceptions. The preparation of detector materials and detector fabrication can present considerable hazards. Some crystal preparation techniques require the use of a toxic substance, eg, Hg, at vapor pressures above 1 MPa (10 atm). Ampul explosions do occur. The electrical circuitry required to operate photodetectors almost always couples to detector devices at low voltages, eg, <10 V; therefore, electrical hazards are minimal. Detectors that are operated at low temperatures are mounted in evacuated containers or Dewars which are partially or nearly completely made of glass. These containers are subject to implosion by improper handling.

BIBLIOGRAPHY

"Photoelectric Cells" in *ECT* 1st ed., Vol. 10, pp. 519–542, by V. K. Zworykin and E. G. Ramberg, Radio Corp. of America; "Phototubes and Photocells" in *ECT* 2nd ed., Vol. 15, pp. 396–443, by C. P. Hadley and co-workers, Radio Corp. of America; "Photodetectors" in *ECT* 3rd ed., Vol. 17, pp. 560–611, by S. R. Borrello, Texas Instruments.

1. R. A. Smith, F. E. Jones, and R. P. Chasmar, *The Detection and Measurement of Infrared Radiation*, Oxford University Press, London, 1968.
2. W. L. Wolfe and G. J. Zissis, eds., *The Infrared Handbook, Rev. Ed.*, Environmental Research Institute of Michigan, Ann Arbor, 1985.
3. P. W. Kruse, L. D. McGlauchlin, and R. B. McQuistan, *Elements of Infrared Technology*, John Wiley and Sons, Inc., New York, 1962.
4. S. M. Ryvkin, *Photoelectric Effects in Semiconductors*, Consultants Bureau, New York, 1964.
5. P. W. Kruse, "The Photon Detection Process," in R. J. Keyes, ed., *Topics in Applied Physics, Optical and Infrared Detectors*, Springer-Verlag, New York, 1980.
6. S. M. Sze, *Physics of Semiconductors*, Wiley-Interscience, New York, 1969.
7. R. D. Hudson, Jr., *Infrared System Engineering*, Wiley-Interscience, New York, 1969.
8. J. A. Jamieson, R. H. McFee, G. N. Plass, R. H. Grube, and R. G. Richards, *Infrared Physics and Engineering*, McGraw-Hill Book Co., Inc., New York, 1963.
9. H. L. Hackforth, *Infrared Radiation*, McGraw-Hill Book Co., New York, 1960.
10. C. S. Williams and O. A. Becklund, *Optics, A Short Course for Engineers and Scientists*, Wiley-Interscience, New York, 1972.
11. M. J. Buckingham, *Noise in Electronic Devices and Systems*, Halsted Press, New York, 1983.
12. S. Borrello and Z. Celik-Butler, *Solid State Electron.* **36**, 407 (1993).
13. R. S. List, *J. Electron.* **22**(8), 1017 (1992).
14. S. R. Borrello, C. G. Roberts, B. H. Breazeale, and G. R. Pruett, *Infrared Phys.* **11**, 225 (1971).
15. C. Hilsum, "Device Physics," in T. S. Moss, ed., *Handbook on Semiconductors*, Vol. 4, North-Holland, New York, 1981.
16. J. H. Longo, D. T. Cheung, A. M. Andrews, C. C. Wang, and J. M. Tracy, *IEEE ED* **25**(2), 213–232 (1978).
17. L. J. M. Lesser and F. L. J. Sangster, "Charge Transfer Devices," in Ref. 15.
18. A. F. Milton, "Charge Transfer Devices for Infrared Imaging," in Ref. 5.
19. E. H. Nicollian and J. R. Brews, *MOS Physics and Technology*, John Wiley and Sons, Inc., New York, 1982, Chapt. 2.
20. C. Y. Wei and co-workers, *IEEE Trans. Electron Dev.* **ED-27**(1), 170 (1970).
21. D. Long, "Photovoltaic and Photoconductive Infrared Detectors," in Ref. 5.
22. M. A. Kinch and S. R. Borrello, *Infrared Phys.* **15**, 111 (1975).
23. E. H. Putley, "Thermal Detectors," in Ref. 5.
24. A. Rogalski, "Selected Papers on Semiconductor Infrared Devices," in J. Thompson, ed., *SPIE Milestone Series*, SPIE Optical Engineering Press, Bellingham, Wash., 1992.
25. W. D. Rogatto, "Electro-Optics Components," in J. S. Accetta and D. L. Shumaker, eds., *The Infrared & Electro-Optical Systems Handbook*, Vol. 3, SPIE Optical Engineering Press, Bellingham, Wash., 1993.
26. W. S. Boyle and G. E. Smith, *BSTJ Briefs* **49**(4), 587 (1970).
27. G. F. Amelio, M. F. Tompsett, and G. E. Smith, *BSTJ Briefs* **49**(4), 593 (1970).
28. J. R. Janesick and co-workers, *Proc. SPIE* **501**, 2 (1984).
29. J. Hynecek, *IEEE Trans. Electron. Devices* **ED-32**(8), 1421 (1985).

30. K. A. Parulski, *IEEE Trans. Electron. Devices* **ED-32**(8), 1381 (1985).

31. Y. Takemura and co-workers, *IEEE Trans. Electron. Devices* **ED-32**(8), 1402 (1985).

32. D. F. Barbe, *Proc. IEEE* **63**, 38 (1975).

33. *DALSA Inc. Scanner News* **7**, 1 (1993).

34. M. M. Blouke, J. R. Janesick, and J. E. Hall, *Optical Eng.* **22**(5), 607 (1983).

35. H. U. Keller and co-workers, *J. Phys. E (Scien. Instr.)* **20**(6), 807 (1987).

36. K. P. Klaasen, M. C. Clary, and J. R. Janesick, *SPIE J.* **331**, 376 (1982).

37. J. V. Sweedler, K. L. Ratzlaff, and M. B. Denton, eds., *Charge-Transfer Devices in Spectroscopy*, VCH Publishers, New York, 1994.

38. T. W. Barnard and co-workers, *Anal. Chem.* **65**, 1231 (1993).

39. J. C. Sutheland and co-workers, *J. Anal. Biochem.* **163**, 446 (1987).

40. P. M. Epperson, R. D. Jalkain, and M. B. Denton, *Anal. Chem.* **61**, 282 (1989).

41. T. W. Barnard, M. I. Crockett, J. C. Ivaldi, S. L. Lundberg, and E. F. Young, *1992 Pittsburgh Conference*, New Orleans, paper no. 990.

42. T. W. Barnard, M. I. Crockett, J. C. Ivaldi, S. L. Lundberg, and E. M. Ziegler, *1992 Pittsburgh Conference*, New Orleans, paper no. 991.

43. R. L. McCreery, in Ref. 37, pp. 227–279.

44. J. Hyncek, *IEEE Trans. Electron Dev.* **ED-38**(5), 1011 (1991).

45. S. Tsuneta and co-workers, *Solar Phys.* **136**(1), 37 (1991).

46. D. N. Burrows and co-workers, in O. H. W. Siegmund, ed., *EUV, X-Ray, and Gamma-Ray Instrumentation for Astronomy III*, *Proc. SPIE* **1743**, 72 (1992).

47. R. D. Thom, T. L. Koch, J. D. Langan, and W. J. Parrish, *IEEE Trans. Electron Devices* **ED-27**(1), 160 (1980).

48. M. A. Kinch, R. A. Chapman, A. Simmons, D. D. Buss, and S. R. Borrello, *Infrared Phys.* **20**, 1 (1980).

49. D. K. Schroder, *Applied Physics Lett.* **25**(12), 747 (1974).

50. K. Y. Han, R. Iyer, M. Hafich, G. Y. Robinson, and D. L. Lile, *Electron. Lett.* **28**(19), 1795 (1992).

51. M. V. Wadsworth, S. R. Borrello, and co-workers, *IEEE Trans. Electron Dev.* **ED 42**(2), 244 (1995).

52. W. Runyon, in M. Grayson, ed., *Encyclopedia of Semiconductor Technology*, John Wiley and Sons, Inc., New York, 1984, p. 809.

53. D. J. Burt, "Fabrication Technology for Charge-Coupled Devices," in M. J. Howes and D. V. Morgan, eds., *Charge Coupled Devices and Systems*, John Wiley and Sons, Inc., New York, 1979.

54. J. Hyncek, *IEEE Trans. Electron. Dev.* **ED-28**(5), 483 (1981).

55. J. R. Janesick and co-workers, in M. M. Blouke and D. Pophal, eds., *Optical Sensors and Electronic Photography*, *Proc. SPIE* **1071**, 153 (1989).

56. R. H. Bube, *Solar Cells*, in Ref. 15.

57. R. H. Bube, *Photoconductivity of Solids*, John Wiley and Sons, Inc., New York, 1960.

58. A. Campion, J. Brown, and W. H. Grizzle, *Surf. Sci.* **115**, L153 (1982).

59. S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed., John Wiley and Sons, Inc., New York, 1981.

60. A. B. Dreeben and R. H. Bube, *J. Electrochem. Soc.* **110**, 456 (1963).

61. J. A. Beun, R. Nitsche, and H. U. Bolsterli, *Physica* **28**, 194 (1962).

62. "Metalorganic Vapor Phase Epitaxy 1992," in the *Proceedings of the Sixth International Conference on Metalorganic Vapor Phase Epitaxy*, Cambridge, Mass., June 1992; 8–11, G. B. Stringfellow and J. J. Coleman, eds., *J. Crystal Growth*, **124**(1–4) (Nov. 1992).

63. E. H. Putley, "Lead Sulphide, Selenide and Telluride," in C. A. Hogarth, ed., *Materials Used in Semiconductor Devices*, Wiley-Interscience, New York, 1965.

64. G. Giroux, *Can. J. Phys.* **41**, 1840 (1963).

65. W. F. Kosonocky, *Proc. SPIE*, **1308**, 2–26 (1990).

66. B. Capone and co-workers, *Opt. Eng.* **18**(5), 535–541 (1979).

67. M. Kimata, M. Dneda, N. Yutani, S. Iwadi, and N. Tsubouchi, *IEEE Solid State Circuits*, **SC 22**(6), 1124–1129 (1987).

68. C. Hilsum and A. C. Rose-Innes, *Semiconducting III-V Compounds*, Pergamon Press, New York, 1961.

69. J. T. Wimmers, R. M. Davis, C. A. Niblack, and D. S. Smith, *Proc. SPIE*, **930**, 125–138 (1988).

70. S. R. Jost, V. F. Meikleham, and T. H. Myers, in R. F. C. Farrow, J. F. Schetzina, and J. T. Cheung, eds., *Symposia Proceedings of the Material Research Society*, Pittsburgh, Pa., 1987, p. 429.

71. C. A. Castro and J. H. Tregilgas, *J. Crystal Growth*, **86**, 138–145 (1988).

72. T. Tang, M. H. Kalisher, A. P. Stevens, and P. E. Herning, in Ref. 70, p. 321.

73. G. N. Pultz, P. W. Norton, E. E. Krueger, and M. B. Reine, *J. Vacuum Sci. Technol. B* **9**(3), 1724–1730 (1991).

74. M. A. Kinch, S. R. Borrello, and A. Simmons, *Infrared Phys.* **17**, 127 (1977).

75. M. Itoh, H. Takigawa, and R. Ueda, *IEEE ED* **27**(1), 150–154 (1980).

76. C. T. Elliott, "Infrared Detectors," in Ref. 15.

77. A. Jyzwikowska, K. Jyzwikowski, and A. Rolgalski, *Infrared Phys.* **31**(6), 543–554 (1991).

78. C. T. Elliott, D. Day, and D. J. Wilson, *Infrared Phys.* **22**, 31–42 (1982).

79. L. O. Bubulac, *J. Crystal Growth*, **86**, 723–734 (1988).

80. J. P. Rosebeck, R. E. Starr, S. L. Price, and K. J. Riley, *J. Appl. Phys.* **53**(9), 6430–6440 (1982).

81. R. B. Bailey and co-workers, *IEEE ED* **38**(5), 1104–1109 (1991).

82. J. P. Rode, *Infrared Phys.* **24**(5), 443–453 (1984).

83. R. E. DeWames, G. M. Williams, J. G. Pasko, and A. H. B. Vanderwyck, *J. Crystal Growth*, **86**, 849–858 (1988).

84. Y. Nemirowsky, D. Rosenfeld, R. Adar, and A. Kornfeld, *J. Vac. Sci. Technol.* **7**(2), 528 (1989).

85. P. R. Norton, *Optical Eng.* **30**(11), 1649–1663 (1991).

86. P. R. Bratt, "Impurity Germanium and Silicon Infrared Detectors," in R. K. Willardson and A. C. Beer, eds., *Semiconductors and Semimetals*, Vol. 12, Academic Press, Inc., New York, 1977.

87. R. D. Nelson, *Optical Eng.* **16**(3), 275–283 (1977).

88. D. H. Pommerrenig, *Proc. SPIE*, **443**, 144–150 (1984).

89. S. B. Stetson, D. B. Reynolds, M. G. Staplebroek, and R. L. Stermer, *Proc. SPIE*, **686**, 48–65 (1986).

90. L. J. Kozlowski and co-workers, *IEEE ED* **38**(5), 1124–1130 (1991).

91. S. D. Gunapala, B. F. Levine, L. Pfeiffer, and K. West, *J. Appl. Phys.* **69**, 6517 (1991).

92. **U.S. Pat. 5,288,649** (Feb. 22, 1994), W. F. Keenan (to Texas Instruments Inc.).

93. R. A. Wood and co-workers, *Proc. GOMAC Conf.* 579 (1991).

Sebastian R. Borrello
Mark V. Wadsworth
Texas Instruments, Inc.



PHOTOELECTROCHEMICAL CELLS.

See PHOTOVOLTAIC CELLS; SOLAR ENERGY.

PHOTOGRAPHY

The unique light-sensing properties of silver halide crystals have been recognized since the 1500s. In spite of many technical advances in nonsilver halide (eg, electronic) technologies, chemically based silver halide systems continue to dominate in the ability to record images (1–3) of superb image quality and archival characteristics. Photochemical reduction in which the silver ion, Ag⁺, in the ionic silver halide crystal is reduced to elemental silver, Ag⁰ (4–7) was first observed by the alchemist Fabricius in 1556. As photochemical reduction continues, elemental silver atoms aggregate and grow into clusters of a colloidal size sufficient to scatter light and produce hue shifts. The science of photography uses this photochemical property of silver halide to form images and record scenes. One of the earliest researchers to produce such a photochemical image was Schultze in 1727 (8). In these experiments, solutions of silver nitrate and white chalk were photochemically reduced to produce metallic silver images when the solutions were exposed through stencils.

The daguerreotype process, prepared by Daguerre in 1837, and the calotype process, produced by Talbot in 1841, were among the first photographic techniques to produce continuous-tone images as reproductions of scenes (9,10). The steps of a typical daguerreotype process include polishing and cleaning a silver-plated copper plate; treating the silver side of the plate with iodine vapors to convert silver into light-sensitive silver iodide [7783-96-2], AgI; exposing the plate through the optics of a camera that projects and focuses a scene on the plate, ie, where light strikes these plates, silver ions are photochemically reduced to silver metal; and treating the exposed plate with mercury. The mercury reacts with silver metal to produce a silver amalgam. White silver amalgam appears in areas of the plate exposed by light; the unexposed areas remain dark, thereby producing a positive image.

Daguerrotype images were produced primarily by the action of light. For the production of satisfactory images, long exposures to light, on the order of minutes, were required because many absorbed photons were necessary to reduce photochemically enough silver ions to silver metal for visible imaging. In 1841, Talbot announced the calotype process, which could reduce exposure times to seconds and produce visible images without dependence on the action of light (11). In the calotype process, the exposure of silver halide produced an invisible latent image that was composed of only trace amounts of reduced silver. This acted as a catalyst for subsequent chemical reduction, ie, a nonphotochemical continuation of the light-initiated reduction process that eventually produced a visible silver image. In addition to being sensitive to much lower light levels, the calotype process was less expensive than the daguerreotype process and provided a negative–positive system that could generate several positive copies from a single negative original. The use of chemical amplification after low level image exposure is a conceptual approach applied in modern photography.

Unexposed silver halide in these early recorded images photolytically darkened upon repeated exposure to light, thus photographic images were not permanent. In 1839 Herschel discovered that unexposed silver halide could be dissolved with sodium thiosulfate and washed away, whereas metallic silver was relatively unaltered by sodium thiosulfate treatments (12). This process, called fixation, along with chemical and spectral sensitization, were necessary to the foundation of modern photography. Sensitization improvements led to the development of high sensitivity available-light photographic systems. Before such sensitivity improvements, high energy illumination was required to give enough exposure intensity.

Without special sensitizing treatments, silver chloride [7783-90-6], AgCl, microcrystals, which do not significantly absorb light having wavelengths greater than 400 nm, are virtually insensitive to visible light. Similarly, silver bromide [7785-23-1], AgBr, and Ag(Br,I) and Ag(Br,Cl,I) microcrystals popularly used in modern photography are effectively insensitive to electromagnetic radiation of wavelengths longer than 500 nm. Photographically these crystals are said to be blue sensitive, but green and red insensitive. Blue sensitivity refers to the intrinsic sensitivity of the silver halide crystals. To reduce the number of blue photons required to produce a developable latent image, ie, a catalytic center, the silver halide crystals are treated with materials called chemical sensitizers, that absorb to the crystal surfaces and may or may not react with them. Chemical sensitizers do not significantly alter the light-absorption properties of the silver halide crystals, but they do alter the efficiency with which the latent image is formed. Sulfur- and gold-containing compounds are among the most popular chemical sensitizers. In 1925 it was demonstrated that in certain samples of gelatin (qv), sulfur could increase the intrinsic sensitivity of silver halide microcrystals (13–15). Gelatin has been used since 1847 as a protective colloid to prevent the silver halide microcrystals from aggregating or coalescing before and after coating on paper, film, or glass supports. The use of gold salts as chemical sensitizers was discovered in 1936 at the Agfa film plant in Germany (16). However the mechanisms by which these salts enhance photographic sensitivity continue to be investigated (17–21).

To achieve photographic sensitivity in the green (500–600 nm) and red (600–700 nm) regions of the visible spectrum, silver halide crystals are spectrally sensitized with dyes (22) (see DYES, SENSITIZING). In spectral sensitization, dye molecules are absorbed to the silver halide surfaces. The transition from glass plates to film bases for supporting light-sensitive emulsions represents the final step that made photography a popular and commercial success.

The first satisfactory photographic film was produced in 1888 when gelatin-dispersed microcrystals of silver halide were coated on celluloid sheets (23). Within a year George Eastman prepared and marketed roll films on a base produced by dissolving nitrocellulose with camphor and amyl acetate in methanol (qv).

A broad range of photographic materials exists in the 1990s including x-ray films (see X-RAY TECHNOLOGY), graphic arts films, microfilms, and complex multilayer coatings for color films that provide low granularity, available-light sensitivity, and rapid access. Many modern color films have more than 15 separate layers. For such films, coatings having as much silver as 10 g m² and as many as one hundred different chemical compounds may be necessary to provide the desired image quality (granularity and sharpness), color reproduction, image permanence, and light sensitivity (see COLOR PHOTOGRAPHY).

The modern preparation of photographic films, papers, and plates begins with the growth of silver halide microcrystals (Fig. 1). Commercially, reaction vessels having capacities as large as 2000 L are used to produce microcrystals (grains) ranging in size from tens of nanometers for microfilms up to micrometers for the highly sensitive crystals required in available-light photography. During the early stage of the grain growth process (precipitation), silver ions from a silver nitrate [7761-88-8] solution and halide ions from an alkali halide salt solution come together in a reaction vessel to form silver halide nuclei. After this stage, the nuclei are available as substrates for continued crystal growth. The crystals are generally precipitated in aqueous environments using a low concentration of added gelatin which, used as a protective colloid, prevents the grains from aggregating into clumps as ionic strength increases. After the precipitation, the counterions, eg, Na⁺, K⁺, and NO₃⁻, are removed by a washing process. The resulting silver halide dispersion in gelatin is then chilled and allowed to gel (emulsion). The emulsion is usually stored in this chilled state for future use.

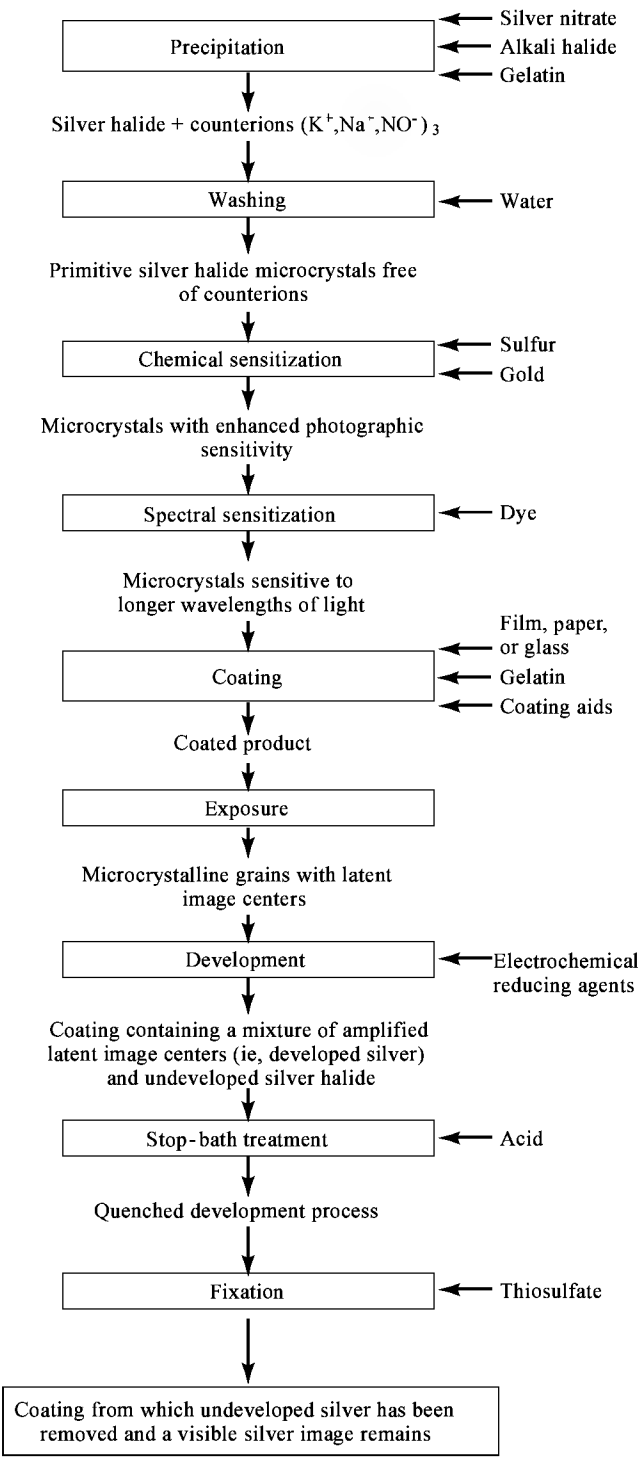


Fig. 1. Flow chart of the photographic process.

Later, the emulsion is melted, the emulsion grains chemically sensitized to enhance their intrinsic sensitivity, and spectrally sensitized using dyes to extend the intrinsic sensitivity into the lower energy regions of the visible spectrum. For certain applications, such as aerial reconnaissance, ecological land surveys, and recording laser diode signals from digital data, emulsions are spectrally sensitized with dyes that sensitize into the infrared region of the spectrum. Figure 2 shows absorptance spectra for AgCl, AgBr, and Ag(Br,I) grains with and without an adsorbed spectral sensitizing dye. Once the grains have been precipitated and chemically and spectrally sensitized, the emulsions are ready for coating on a support (Fig. 3). The choice of support depends on usage requirements. Paper, glass (qv), and polymeric films are most popular because of the dimensional stability, chemical inertness, archival properties, flexibility, and convenience they exhibit. Desired intergrain separations are achieved by coating silver and gelatin in appropriate concentration ratios. Before coating, spreading agents (surfactants) and high molecular weight polymers are added to the emulsions to facilitate coating operations, antifoggants are added to improve the signal-to-noise performance of the photographic film, hardeners are added to produce thermally stable gelatin matrices, emulsion stabilizers are added to extend shelf-life properties, and in the case of color photography (qv), organic compounds called couplers or dye-release materials are added to allow for the production of colored images.

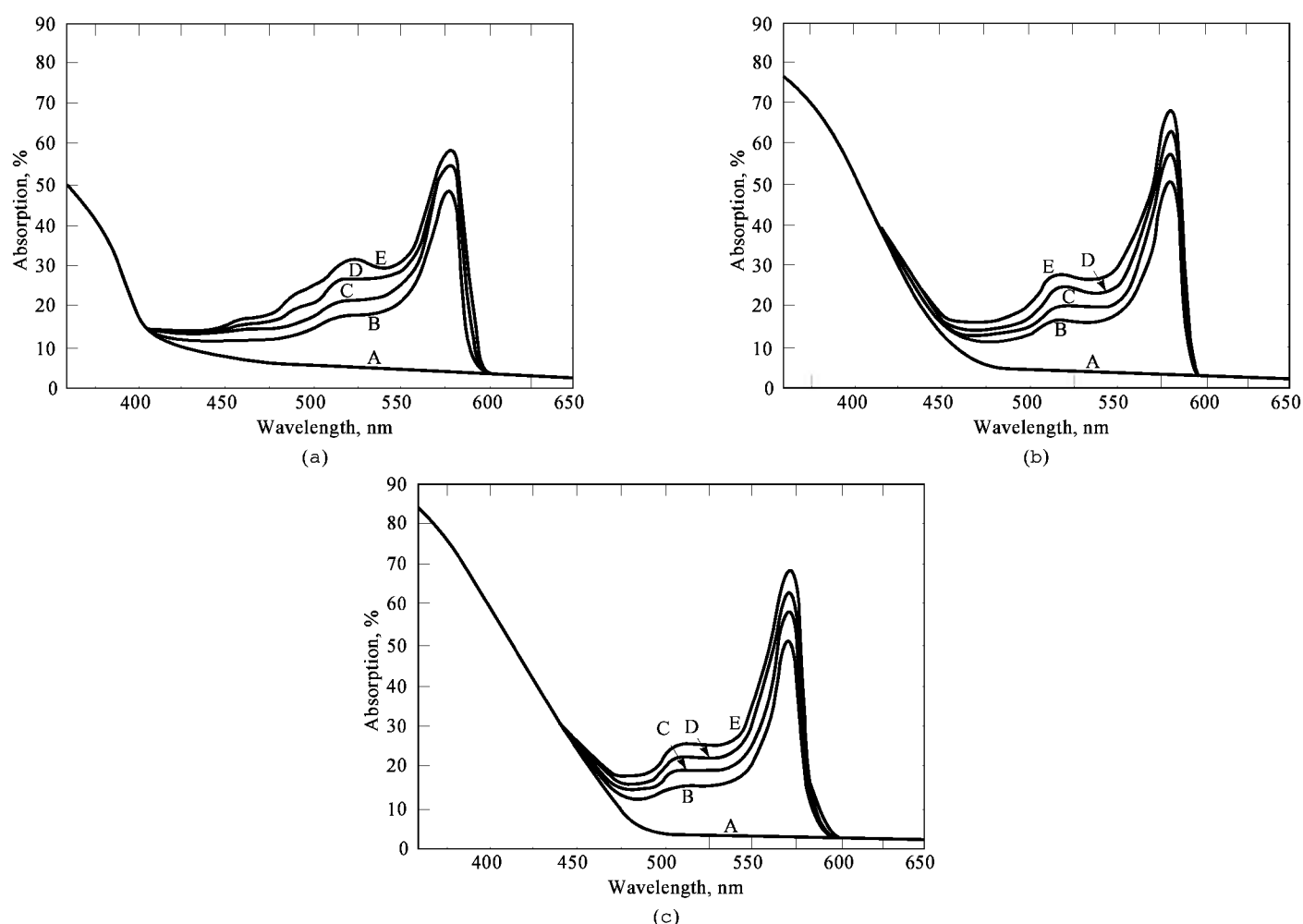


Fig. 2. The effect of a sensitizing dye (1,1'-diethyl-2,2'-cyanine iodide) on the absorption spectra of (a) AgCl, (b) AgBr, and (c) Ag(Br,I) crystals (iodide content $\approx 2.5\%$). The dye levels, ie, A = 0, B = 3.3×10^{-4} , C = 4.4×10^{-4} , D = 5.5×10^{-4} , and E = 6.6×10^{-4} dye mol/mol of silver, correspond to 0.5–1.0 dye mol/nm² of grain surface.

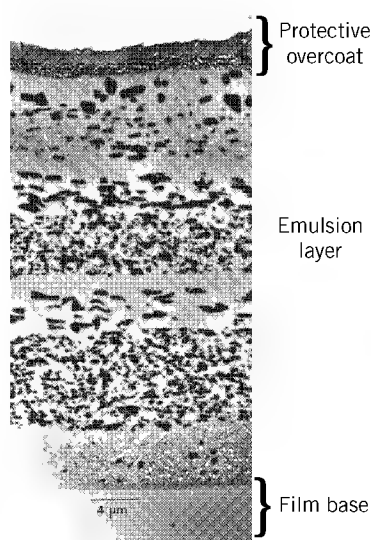


Fig. 3. Cross-section photomicrograph of a color-negative product showing the film base, the emulsion layer (the black specks are microcrystalline silver halide grains), and a protective overcoat. The emulsion layer and overcoat are $\sim 3.5 \times 10^{-5}$ m thick.

Once coated, the photographic material is in an appropriate form to be exposed. The optics of the camera focus an image of a given scene onto the emulsion grains. When an exposing photon is absorbed by the silver halide grain, a mobile electron and a mobile hole are produced within the microcrystalline grain. The photoelectrons thus produced are available to reduce silver ions electrochemically to elemental silver. As the exposure continues, a stable cluster of silver atoms (a latent image) is eventually formed, and the signal is recorded. The number of atoms composing the latent-image center is below the detection limits of conventional light microscopy, yet the cluster is large enough to function as a catalytic center for the electrochemical reduction of more silver ions by developer molecules. Continuing advances in atomic force microscopy may eventually allow for detailed detection and analysis of sensitivity and latent-image centers (24,25).

Development is an amplification stage following exposure in which the relatively small number of silver atoms comprising the latent-image center is magnified by factors as large as 10^{10} . Development can occur within seconds after exposure, as in instant photography (26–30) (see COLOR PHOTOGRAPHY, INSTANT), or it can be months or years after exposure, as in conventional photography. When amplification occurs in the exposed regions of a negative film, a signal composed of elemental silver is produced; however, when amplification occurs in the unexposed regions of such a film, the relative signal intensity is diminished. Such unwanted amplification in the unexposed regions (or in the exposed regions of a reversal film) is called fog or D-min (minimum density). In a conventional black-and-white film, the developed silver produces visual darkening by scattering and absorbing light. This visual darkening is measured in units of optical density, the logarithm of the ratio of the incident-to-transmitted light for a transparent support, or the ratio of the

incident-to-reflected light for a reflection support. After development the coatings are transported to a bath containing thiosulfate, where the remaining undeveloped silver halide is removed by fixation. During fixation, water-soluble complexes are formed between silver and thiosulfate ions, and are later washed from the coatings. The silver may be recovered for future use. After the undeveloped silver ions are removed, a stable silver record of the original scene remains. In instant color films there is no fixation.

The extent of silver-ion reduction during development depends on the development time, the composition and temperature of the developer solution, and the level of the original exposure to light. The relation between the extent of development or amplification and the exposure is often expressed graphically in the form of a characteristic plot of the optical density, D , as a function of the logarithm of the exposure, $\log H$ (31–33). Figure 4 is an example of a negative-working characteristic curve. In the D-min region (fog), the exposure is below the detection limits of the system. As the exposure increases, the density increases up to the saturation–exposure region. The slope of the characteristic curve in the mid-scale region is defined as the contrast of the photographic response, and the exposure range from toe to shoulder measures the latitude of the response; both are important in determining how well a scene is reproduced. In addition to contrast and latitude, the image also is evaluated in terms of sharpness, granularity, and color reproduction (34,35).

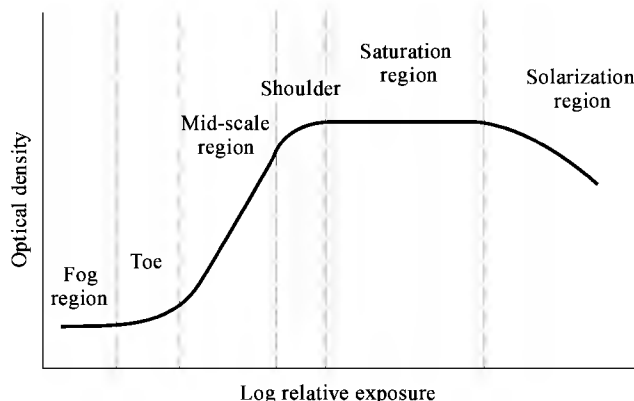


Fig. 4. A characteristic D - $\log H$ curve.

The Photographic Crystal

The preparation of light-sensitive photographic materials begins with the precipitation of silver halide grains. Various process control parameters are rigorously regulated during crystal growth to achieve the desired grain morphologies, size-frequency distributions, solid-state properties, light sensitivity, and catalytic activity (developability). In addition to the use of mechanical process control parameters, eg, flow and mixing rates, various chemicals are added during precipitation to control silver halide growth rates, ripening characteristics, stability, and even light sensitivity. Common to all precipitations is the controlled mixing of solutions of a halide, such as an alkali halide salt, and a silver salt, usually silver nitrate, in the presence of a peptizing agent. Gelatin is one of the best peptizing agents known; however, others have been used (36). During mixing, a chemical reaction occurs and a suspended solid phase, in the form of microscopic silver halide crystals, separates from the liquid phase. As the reactants are added to the reaction vessel, the concentrations of the soluble counterions (alkali cations and nitrate anions) increase. If the ionic strength of the solution in the reaction vessel is sufficiently high, the double-layer repulsion force (37) between grains can be reduced to less than the van der Waals attraction force, a condition that encourages the coagulation of the silver halide particles. The presence of a peptizing agent that adsorbs to the grain surfaces but does not inhibit continued growth is essential to prevent coagulation and to maintain a uniform dispersion of microcrystalline grains.

Techniques of Crystal Growth. Two commonly used precipitation techniques for crystal growth are the single-jet and the double-jet methods (38–40). In the single-jet approach an aqueous solution of the halide salt and peptizing agent is in the reaction vessel at thermal equilibrium before precipitation begins. The chemical reaction is initiated by the controlled addition of a silver nitrate solution. The silver ion activity increases monotonically throughout this process, whereas the ionic strength starts and remains at a high value. In double-jet precipitations, the halide and silver salts are added simultaneously to the reaction vessel, allowing the operator to control the silver ion activity as the reaction progresses. The ionic strength increases monotonically in this method. In addition to the use of gelatin, other approaches have been developed either to control or overcome the adverse effects of high ionic strengths during crystal growth. High amplitude mechanical vibrations (41) and high dilutions (42) during the precipitation are two procedures that have reduced or even eliminated the need for peptizing agents. Triple-jet precipitations, in which halide ions, silver ions, and silver halide seed crystals are simultaneously added to the reactor, have been described in the literature (43) and offer an approach for achieving grain size distributions not achievable by the other methods.

The silver ion activity during precipitation is an important parameter in determining the morphology of the resulting silver halide crystals. In general, high silver ion activities favor cubic structures bound by $\langle 100 \rangle$ crystallographic surfaces; lower activities promote the formation of octahedral grains bound by $\langle 111 \rangle$ surfaces. At very low (10^{-9} – 10^{-11} M) silver ion activities, crystals having single and multiple twin planes are produced. When double parallel twin planes are generated in a given AgBr crystal, thin tabular crystal morphologies can be achieved. pAg , defined as $-\log [Ag^+]$, is often used as a parameter related to silver ion activity, ie, high pAg corresponds to low silver ion activity. The pAg in a reaction vessel determines the number and type of silver halide complexes, eg, Ag_2Br^+ , $AgBr_2^-$, $AgBr_3^{2-}$, that are present (44) and thus strongly affects the silver halide solubility, which strongly influences crystal growth rates. Thus the solution pAg in an emulsion precipitation reactor is an important control parameter, affecting both crystal size and morphology. Because of this silver ion activity effect, single-jet precipitations, which have low initial silver ion activities, generally produce $\langle 111 \rangle$ crystallographic surfaces.

Various organic compounds can be added to the reaction vessel during crystal growth to modify or control the resulting crystalline morphology. For example, silver chloride, which under most conditions grows with cubic structure (45), can be prepared with octahedral and dodecahedral morphologies if habit modifiers are used (46). Figure 5 shows electron micrographs of cubic, octahedral, and tabular silver bromide grains. Figure 6 shows the atomic arrangement within a silver halide unit cell and orientation of three important crystallographic surfaces: silver halide cubes are bounded by $\langle 100 \rangle$ surfaces, dodecahedra are bounded by $\langle 110 \rangle$ surfaces, and octahedra are bounded by $\langle 111 \rangle$ surfaces. The arrangement of lattice ions and the associated interlattice site distances (47–49) on perfect surfaces are exhibited in Figure 7. Newly developed analytical techniques, such as surface extended x-ray absorption fine structure (sexafs) and atomic force microscopy (afm), suggest that the structure and interionic separations on the surface of silver halide microcrystals are the same as in the bulk for $\langle 100 \rangle$ AgBr surfaces, but that some reconstruction may occur on $\langle 111 \rangle$ surfaces because of electrostatic effects (50,51). In spite of the presence of gelatin during the growth of silver halide microcrystals, and coulombic forces driving reconstruction, the surfaces can be remarkably smooth. Atomically smooth $\langle 100 \rangle$ AgBr surfaces for thin-film epitaxial vapor deposition on cleaved NaCl substrates have been observed using afm (52) (see THINFILMS).

The ultimate size–frequency distribution of the emulsion grains depends on the rate of addition of reactants, halide type, temperature, and presence of growth modifiers or ripeners. Ripeners are compounds that form water-soluble silver salts or complexes and preferentially dissolve the smallest grains from a given population of crystal sizes to enhance growth in the larger grains. In addition to influencing grain growth kinetics, ripeners can also affect the surface structure of crystals (53). Ammonia (54,55), sodium thiosulfate (56), sodium thiocyanate (57), certain thioethers, and thiourea (58) are effective

ripening agents. Ripeners can be added to the reaction vessel after precipitation to produce post-precipitation digestion effects. Alternatively, ripeners can be introduced into the vessel during or even before the onset of crystal growth.

Silver halide crystals grown from a single halide type have received considerable attention in basic studies. Some studies have focused on the growth of AgBr (59–62) and others on AgCl (63). Although analyses of the growth processes for crystals of mixed halide content, eg, Ag(Br,I) and Ag(Cl,Br,I), have not been as extensively reported in the literature, mixed halide grains of various compositions commonly are used in practical photographic materials. In the preparation of mixed halide crystals, the various halide types may be added to the reaction vessel either simultaneously or sequentially. In the latter case, the most soluble silver halide crystal, ie, AgCl, is sometimes grown first, and through subsequent halide additions the silver halide is converted partly to a silver halide of lower solubility, eg, AgBr. Photographic (64–66), electron-microscopic (67), and x-ray diffraction studies (68–70) suggest that this conversion involves a surprisingly efficient intragranular anionic diffusion. If the least soluble silver halide crystals are grown first, epitaxial growth can be achieved. In epitaxial growth the second phase is crystallographically oriented on the substrate phase (71). For example, AgCl growth on preformed β -AgI microcrystals has been reported (72). The two phases (AgCl and AgI) of the epitaxial crystals are in electrical contact. In fact, mobile photogenerated electrons produced in the AgI phase have been shown to migrate into the AgCl phase.

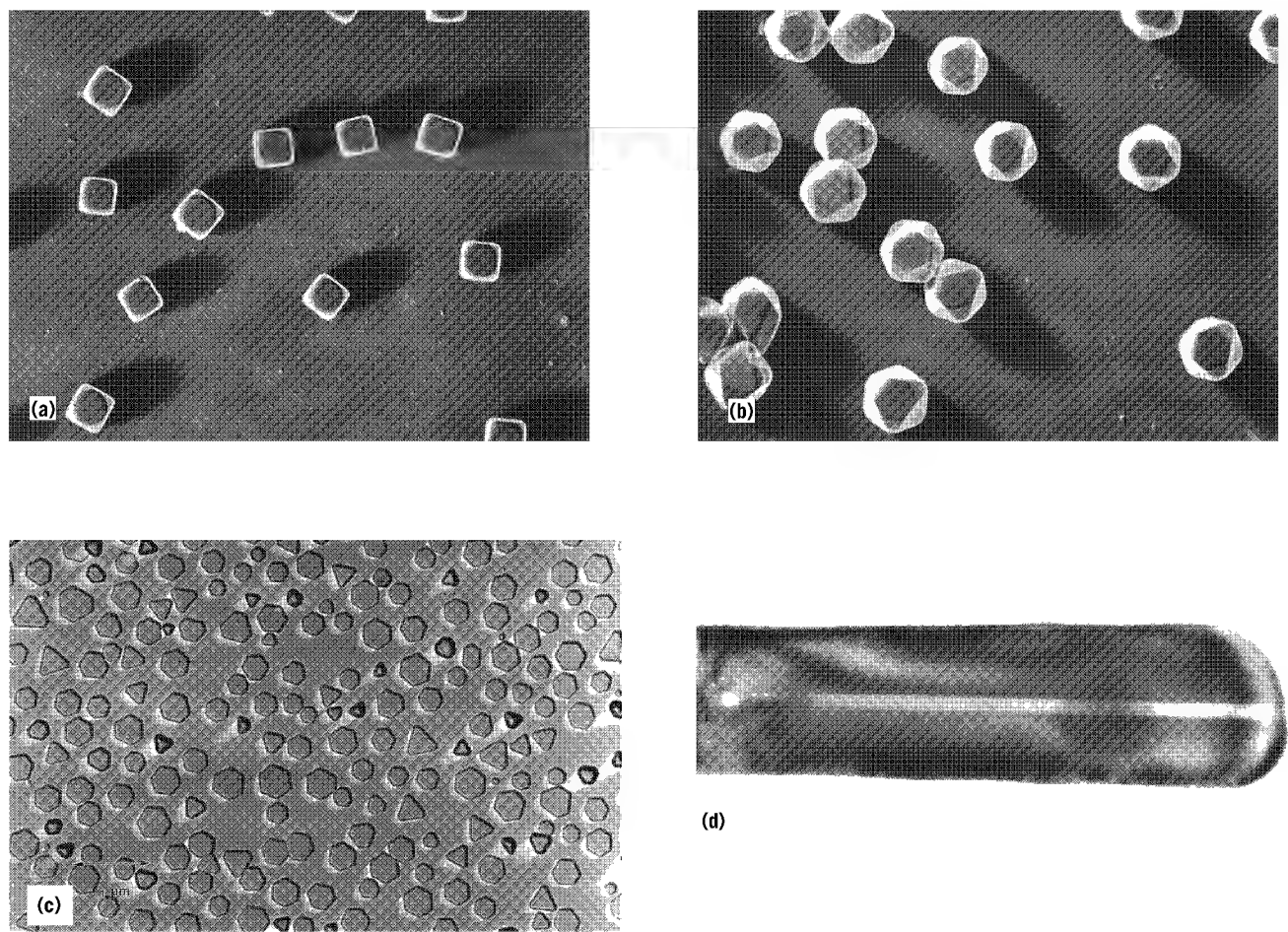


Fig. 5. Silver halide grain morphologies for (a) cubic, precipitated in an environment having a silver ion concentration, $[Ag^+]$, of $ca\ 2.5 \times 10^{-5}$ mol L; (b) octahedral, $ca\ 6.0 \times 10^{-9}$ mol L; and (c) tabular microcrystals, $ca\ 1.0 \times 10^{-10}$ mol L. A cross section of a tabular grain revealing double parallel twin planes is shown in (d).

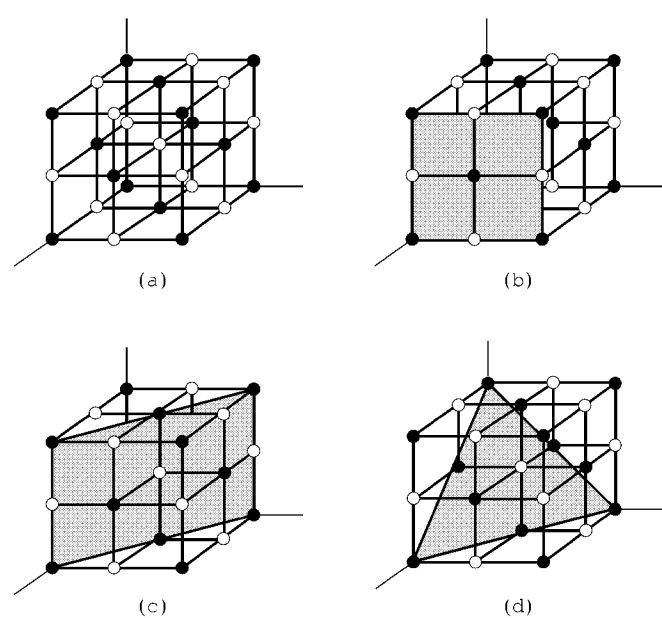


Fig. 6. Unit cell for silver halide where shaded areas represent crystallographic planes, (○) halide ions, and (●) silver ions. (a) Complete cell; (b) showing a $\langle 100 \rangle$ surface; (c), a $\langle 110 \rangle$ surface; and (d), a $\langle 111 \rangle$ surface.

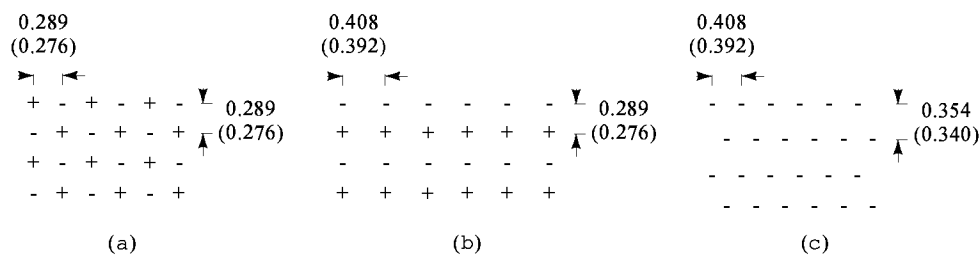


Fig. 7. Ionic arrangements on different (a) $\langle 100 \rangle$, (b) $\langle 110 \rangle$, and (c) $\langle 111 \rangle$ crystallographic surfaces. The interionic separations are given in nanometers for AgBr and in parentheses for AgCl. Values are larger for AgBr because of the relative ionic radii of bromide and chloride.

The growth of silver halide crystals can be viewed as a collection of subprocesses, including nucleation, dissolution, ion diffusion, surface integration of ions, aggregation (73), and growth, which occur concurrently and or sequentially. According to the Gibbs-Thomson equation, sometimes referred to as the Kelvin equation, a small spherical particle having an associated small radius of curvature, r , has a higher solubility, C_r , than the bulk thermodynamic equilibrium solubility, C_∞ , measured over a flat macrosurface. The solubility, C_r , of spherical crystals is given by

$$C_r = C_\infty \exp\{2\sigma V_m / rRT\} \tag{1}$$

where σ is the partial surface free energy ($\partial G / \partial \text{area}$) of the crystal and is by analogy related to the surface tension at a liquid–vapor interface, V_m is the crystal molar volume, R is the gas constant, and T is the temperature. Accordingly, small nuclei having small values of r are relatively soluble and dissolve if the solution concentration is not greater than C_r . In the beginning of a double-jet precipitation the concentrations of silver and halide ions continue to increase until a condition of sufficiently high supersaturation is achieved, the adverse free energy effects described by equation 1 are overcome, and nucleation occurs. The reaction to form a silver halide crystal is initiated by nucleation (74), which is the initial formation of a solid silver halide phase. Once the nuclei are formed, a substrate or surface is available for continued growth, and the growth stage is initiated. The transition between nucleation and growth does not necessarily occur at a specific time during the precipitation or even over a specific time range. Indeed, under certain conditions, eg, in localized regions of supersaturation, nucleation may continue throughout precipitation (59,75). The surface integration process does not appear to be rate determining for double-jet precipitations of AgBr crystals (76) but may play an important role during the formation of twinned planes which are essential for the production of tabular crystals (77), as shown in Figure 5c. Growth models based on diffusion, the Gibbs-Thomson effect, and mass balance have been developed that can predict the number of nuclei formed and the ultimate grain sizes (78,79). Using such a model the number of crystals, Z , grown is given by

$$Z = (FRT(a / r_c - 1)) / 8\pi\sigma V_m DC_\infty \tag{2}$$

where F is the molar addition rate of the reactants (Ag^+ and Br^-); D is the diffusion constant for these ions; C_∞ is the equilibrium solubility which is a function of $p\text{Ag}$, T , and halide type; a is the average effective crystal radius; and r_c is the critical crystal size (see eq. 1) which depends on the supersaturation condition, C / C_∞ , and therefore on the molar flow rate to which this ratio is proportional. Equation 2 is consistent with the observation that larger crystals, ie, fewer nuclei, are achieved with low molar addition rates and high solubility conditions.

In addition to nucleation, growth, halide conversion, and epitaxy, the Ostwald ripening crystallization process must be considered (59,60,80–82). Ostwald ripening occurs when there is a range of grain sizes within the reaction vessel. From equation 1, the solubility of the smaller grains is greater than that of the larger grains and thus the small grains tend to dissolve and contribute to accelerated growth of the larger grains. This dependency of solubility on size revealed by equation 1 is an important property contributing to the ability to produce relatively monodispersed size-frequency distributions. In double-jet precipitations, nucleation has been observed to occur continuously throughout the growth process, yet the total number of crystals does not change after the first few minutes (59,78) because the newly formed nuclei dissolve and redeposit on existing crystals according to the size-dependent driving forces suggested in equation 1. Ostwald ripening can be divided into at least three mechanistic steps: dissolution of small grains, ionic diffusion through the aqueous phase, and redeposition of ions on large grains. The rate-limiting step depends on environmental conditions. However, in general, ripening is promoted by temperature increases and addition of ripening agents which complex with silver forming water-soluble silver salts.

Since the mid-1980s, tabular crystals (Fig. 5c) have become increasingly popular in color as well as black-and-white photographic products. The popularity is a consequence of the relatively high surface-to-volume ratios attainable with the tabular morphologies. For tabular grains having sensitizing dyes adsorbed to their surfaces, the high surface-to-volume ratios render them particularly sensitive to spectral regions of the visible spectrum. Thus, high signal-to-noise (ie, photographic light sensitivity to granularity) signals can be achieved.

It is generally believed that tabular crystals are produced when faults in the buildup of successive lattice layers occur during early stages of crystal growth and parallel twin planes are produced (83–86) (Fig. 5d). Studies indicate that tabular crystals having high aspect ratios can be produced even when the crystals have only a single twin plane (87). Other research has focused on diffusion-controlled lateral growth (88–90) which must predominate over thickness growth to preserve tabular morphology. These models can predict the rate of change in grain diameters and how these rates depend on diffusion coefficients, supersaturation, and time.

Impurity Incorporation. For some applications, inorganic impurities are intentionally added to emulsions during precipitation to achieve certain desired photographic responses. These impurities or dopants are usually incorporated into the silver halide grains at very low (ie, in the ppm or ppb range) concentrations. In spite of these low concentrations, dopants can have significant effects on the solid-state properties of the grains as well as on light sensitivity, contrast, and developability. By the use of photoluminescence (91–93), dielectric loss (94–96), spectrophotometry (97), conductivity (97,98), diffusion (99,100), ionic thermocurrent (101–103), electron spin resonance (104,105), and Auger spectroscopic (106) measurements, dopants have been found to influence the ionic and electronic properties, light absorption, and catalytic properties of silver halide crystals. For example, polyvalent cationic dopants can reduce the interstitial silver ion concentration, influence trapping phenomena associated with mobile conduction band photoelectrons and mobile valence band photoholes, and alter electron–hole recombination probabilities. Some key elements capable of providing these solid-state effects include Ir (107–109), Pb (110–111), Zn (112), Ce (113), Cd (114), Rh (115–118), Cu (119), Fe (119), Os (120,121), and Pd (122). Many of these elements, even in the cationic state, are present as trace level impurities in a variety of raw materials and in the general environment. Because these materials all can have substantial effects on the fundamental processes associated with the photographic sensitivity of emulsion grains, great caution must be exercised in the manufacture of photographic products.

Emulsion Washing and Concentrating Procedures. After crystal growth is completed, the resulting photographic emulsion is a dispersion of silver halide microcrystalline grains in an aqueous gelatin phase. Precipitation by-products such as counterions, ripeners, and others are also present in solution. If these by-products are not removed, adverse crystallization may occur when the emulsion is coated on a support and dried, and certain by-products may interfere with subsequent chemical and spectral sensitization. When photographic materials are coated on a paper support, some of the by-products are adsorbed on the fiber matrix of the paper, which can minimize or eliminate the adverse effects of reaction by-products. However, for many photographic products, such as those coated on film, glass, and water-impermeable resin-treated paper, the adsorbent properties of the substrate are not sufficient, and the amount of water present after crystal growth is often inconveniently high for coating operations.

Several washing procedures have been devised to remove water and water-soluble by-products. The most common of these procedures are noodle

washing, flocculation washing, and ultrafiltration (qv). In noodle washing, additional gelatin is added to the emulsion after precipitation causing the emulsion to solidify into a jelly upon cooling. In the jellied state the emulsion can be fragmented into noodles that are washed with circulating water. Noodle washing is often inconvenient because of high gelatin content and the length of time required to leach by-products.

Flocculation washing requires less time for efficient by-product removal, does not require addition of extra gelatin, and provides a more concentrated emulsion through removal of the water used during precipitation. In flocculation, precipitation gelatin is made to coagulate either by adjusting the pH to the isoelectric point of the protective gelatin colloid (123–125) or by adding salts, eg, inorganic sulfates, which promote coalescence by increasing the ionic strength (126). When the gelatin floccules settle to the bottom of the reaction vessel, they carry the silver halide grains with them. The water-soluble by-products and supernatant that are left behind can be removed by decantation. A sequence of redispersal with water, flocculation, and decantation can be repeated until the silver halide emulsion has the desired purity. Variants of this process have been developed to accelerate settling rates of the floccules (127) and to improve the photographic response of the washed emulsions (128) (see FLOCCULATING AGENTS).

Ultrafiltration is the most modern method for washing and concentrating photographic emulsions (129). In this method, the emulsion is pumped into contact with a semipermeable membrane. A maintained pressure drop across the membrane drives water, counterions, and other low molecular weight precipitation by-products through membrane pores of an optimized size, leaving behind the silver halide crystals and gelatin in a purified and concentrated state suitable for subsequent sensitizing and coating operations.

Crystal Properties. Silver halide crystals have been the photosensitive materials of choice in most photographic materials since the pioneering efforts of Daguerre, Niépce, Talbot, and Herschel in the early 1800s. This preeminence of silver halide in photography is the result of a unique collection of physical and chemical properties. The low aqueous solubility and range of silver halide complexes are important during crystal growth. Solubility products, K_{sp} , for the silver halides are as follows:

Temperature, °C	AgCl, M^2	AgBr, M^2	AgI, M^2
25°	1.77×10^{-10}	4.85×10^{-13}	8.28×10^{-17}
50°	1.30×10^{-9}	6.45×10^{-12}	2.49×10^{-15}

Optical absorption properties, mobility of both ionic and electronic charge carriers, and electronic energy structure are all essential features of silver halide grains during latent-image formation. The catalytic properties of the small silver specks constituting the latent image are required during development. The dissolution characteristics of silver halide crystals permit stabilization during fixation. Because of the outer electronic shell structures for silver and the halogens, ie, Ag, $5s^24d^9$; Cl, $3s^23p^5$; Br, $4s^24p^5$; and I, $5s^25p^5$, the silver halides are often referred to as ionic crystals. Silver has a relatively ionizable electron that it sacrifices to a halogen atom which has a relatively large electron affinity. However, the large dielectric constants, high insolubility in water, and high mobility of interstitial silver ions are not predictable based on an ionic crystal model. These values are all high compared to the values for the alkali halides (130,131).

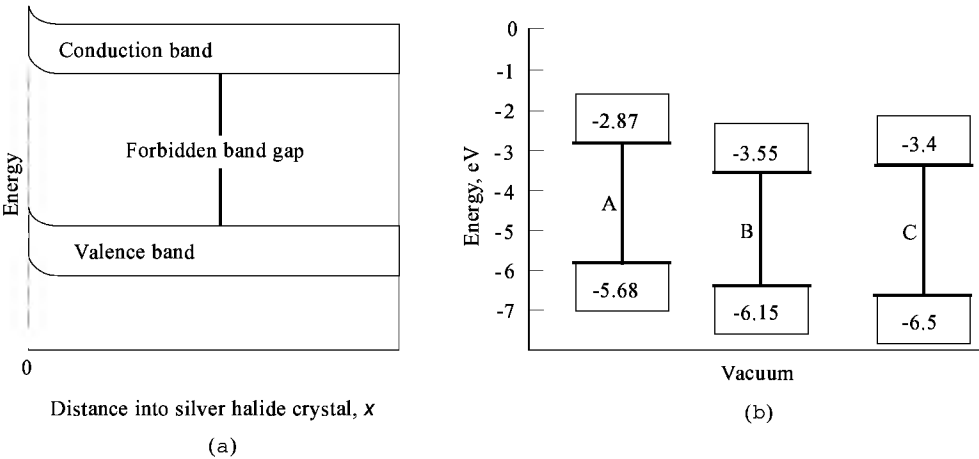


Fig. 8. (a) Energy levels for the band model of silver halide crystals. The band bending at the surface (---) is exaggerated. The extent of bending is at most 0.1 of the band gap. (b) Energy positions relative to vacuum of the conduction and valence bands within crystal bulk. The band gap for β -AgI. A 2.81 eV (4.52×10^{-19} J); for AgBr, B 2.6 eV (ca 4.17×10^{-19} J); and for AgCl, C 3.1 eV (4.99×10^{-19} J) (133,134). See text.

Figure 8a shows a schematic representation of the electronic energy structure of silver bromide. Because of the periodicity of the lattice positions in silver halide crystals, quantum-mechanical analysis dictates an energy-band scheme (132). In this scheme bands of allowed and closely packed electronic energy levels are separated by forbidden energy gaps. The highest energy band in which all of the electronic energy states are occupied is the valence band; the band that is energetically directly above the valence band is the conduction band. In the absence of external excitation energy, the electronic energy states comprising the silver halide conduction band are unoccupied. During exposures of sufficient energy, a photon is absorbed and simultaneously an electron is promoted from the valence band into the conduction band. Such an intrinsic absorption produces a positive mobile hole in the valence band and a mobile electron in the conduction band. The electron and hole thus formed are free to move through the crystal independently and contribute to a photoinduced electrical conductivity. The forbidden gap is temperature dependent because of phonon-assisted transitions which reduce the photon energy required to effect promotion of an electron into the conduction band. At 298 K, the forbidden energy gap for silver bromide is $\sim 4.2 \times 10^{-19}$ J (2.6 eV) (133–136). Therefore, to excite an electron from the valence band into the conduction band, a minimum photon energy of 4.2×10^{-19} J is required. This energy corresponds to the intrinsic silver bromide absorption edge near 475 nm (see Fig. 2).

The energy positions for the conduction and valence bands in bulk β -AgI, AgBr, and AgCl are provided relative to vacuum in Figure 8b. The energy gap for silver chloride is significantly greater than the gap for either silver iodide or silver bromide, as can be correlated with the long-wavelength absorption edges shown in Figure 2. Because these band gap energies are large compared to thermal energy, silver halide crystals do not exhibit dark conductivity resulting from electronic motion. In silver halide crystals, electronic conductivity must be photoinduced. The photoinduced charge carriers produced upon exposure to light initiate the latent image formation in photographic materials. Once formed, the mobile electronic charge carriers do not necessarily remain free, but rather are likely to be trapped at localized sites within the crystal interior or on the crystal surface. The concentration and nature of the trapping centers in nominally pure silver halide crystals, although not completely understood, have been extensively studied (19,137–141). An electron trap is an unoccupied electronic energy state that is energetically below the bottom edge of the conduction band. On the other hand, a hole trap is an occupied electronic energy state above the top edge of the valence band. These traps may result from impurities; surface defects, eg, kink sites; interior defects, eg, jogs on dislocation lines; or intentional dopant additions during or after crystal growth. The depth, cross section, and concentration of such traps determine in part the lifetimes and mobilities of the photoinduced charge carriers.

The ultimate trapping site for a photoelectron is influenced by the high dielectric constant of silver halide (ca 12.5, 11.15, and 7.15 for AgBr, AgCl, and β -AgI, respectively), the negative surface charge, and relative trap depths. Interior traps located at point defects on dislocation lines are probably not as

deep as surface traps at positive kink sites because of the high dielectric constant of silver halide. However, the negative surface charge of the microcrystals appears to repel the photoelectrons and thereby enhances the probability of internal trapping (142). The relative photographic sensitivity of the interior of silver halide microcrystalline grains is consistent with such an effect from a negative surface charge. Many emulsion grains have been designed to shuttle photoelectrons and photoholes in opposite directions to avoid inefficiencies in photographic sensitivity associated with electron–hole recombination (143–147). By creating relatively iodide-rich regions during crystal growth, a driving force that directs photocarrier motion is established within the microcrystal. It is generally believed (134) that both the valence and conduction band edges are higher in iodide-rich regions (Fig. 8b); thus electrons in the conduction band attempt to minimize their energy by moving toward the iodide-poor regions, and holes in the valence band associated with the iodide-depleted regions migrate in the opposite direction. The band positions relative to vacuum in Figure 8b may shift somewhat due to ionic migration of interstitial silver ions and vacancies when two iodide phases are brought into electrical thermal contact (134,148,149).

Although there are no mobile electrons in the dark, silver halide crystals do have a dark conductivity that results from ionic motion. Thermally activated ionic mobility at room temperature contributes to latent-image formation during exposure to light. The ionic motion can result from one of two sources in pure silver halide crystals: movement of silver ions in interstitial positions and the motion of silver ion vacancies. A vacancy is created when a silver ion lattice position becomes unoccupied. Vacancy motion results from successive hopping in which lattice silver ions and adjacent vacancies exchange positions. Of the two ionic charge carriers, the interstitial ions are the more mobile and usually considered the main ionic actors in latent image formation. The concentrations of interstitial ions and vacancies are governed by mass balance equations. A dynamic equilibrium (the Frenkel equilibrium) is established in silver halides such that the mathematical product of the two activities is constant at any given temperature (150).

$$K_f(T) \quad [Ag_i^+] \left[V_{Ag^+} \right]$$

Therefore, any perturbation acting on the crystal to increase or decrease the concentration of one species must have the opposite effect on the concentration of the other species (142).

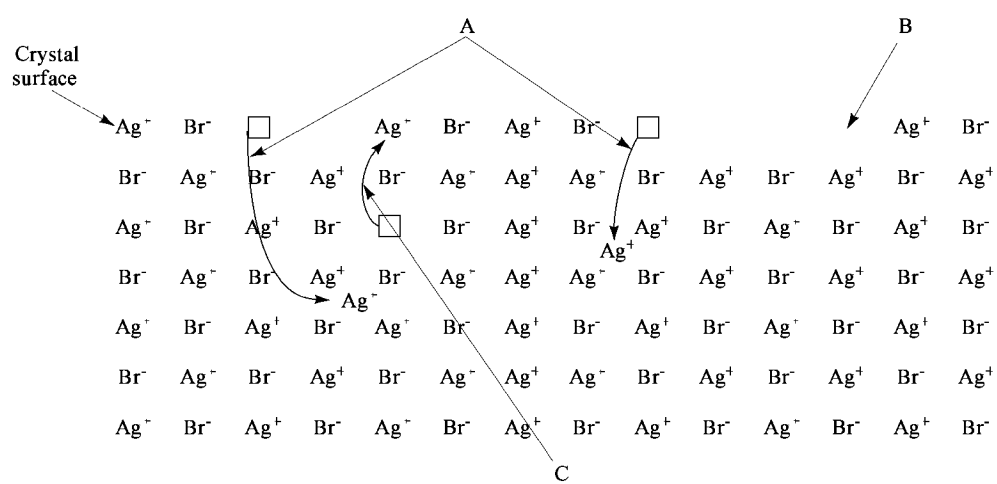


Fig. 9. Schematic of a two-dimensional cross section of an AgBr emulsion grain showing the surface and formation of various point defects: A, processes forming negative kink sites and interstitial silver ions; B, positive kink site; and C, process forming a silver ion vacancy at a lattice position and positive kink site on the crystal surface.

In addition to being affected by polyvalent impurities, the interstitial ion concentration is also influenced by other factors, eg, crystal environment, temperature, species adsorbed to the crystal surface, halide ion, crystallographic morphology (141), and surface space charge (151–154). In fact, the concentration of interstitial silver ions is considerably higher in microcrystals than in bulk silver halide, presumably because of the high surface-to-volume ratios which enhance surface space charge effects. These surface effects occur because the Gibbs free energy required to transport a surface silver ion to an interstitial position is less than the energy necessary for moving a lattice silver atom to the surface, leaving behind an internal silver ion vacancy (155) (Fig. 9). Therefore, thermal equilibrium results in negatively charged surfaces, subsurface space charge regions with high interstitial silver ion concentrations relative to the bulk, and the associated band bending (see Fig. 8a). In spite of their high surface-to-volume ratios, tabular grains have lower ionic conductivities than do cubic crystals and both tabular and cubic microcrystals have significantly lower ionic conductivities than the octahedral crystals. This is presumed to be the result of interstitial silver ion localization at twin planes (141).

Response Enhancement

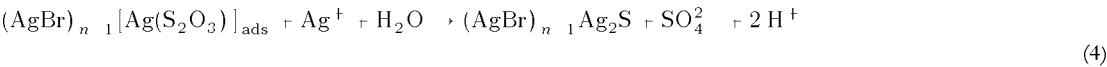
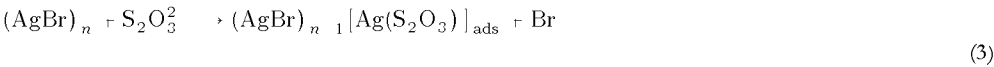
Chemical Sensitization. After the photographic microcrystals are precipitated but before they are coated on a support, the crystals are treated to enhance their sensitivity to light. Chemical sensitization is a process which improves that ability of the emulsion grains to use the absorbed photons, independent of the wavelength. Various methods of post-precipitation chemical sensitization have been developed to reduce the number of photons required to produce a developable latent-image center.

During exposure to light, photons are absorbed by silver halide microcrystalline grains and concurrently photoelectrons and photoholes are formed. These mobile charge carriers then interact with interstitial silver ions in a chemical sequence of mechanistic steps to produce centers of metallic silver called latent-image centers. There is some disagreement about the order of the steps and the identity of the rate-determining step in latent-image formation (130,142,156–167), but the basic scheme for the growth of the silver center and various sources of inefficiencies are widely accepted. The overall desired result of the sum of the various mechanistic steps is the reaction of photoelectrons with silver ions to form localized clusters of silver atoms. There are at least two well-defined loss processes that interfere with this chemical reduction of silver ions: recombination of photoelectrons with photoholes, and reoxidation of silver atoms upon attack by photoholes. In a practical sense, the topography (location) of latent-image center formation on or within the microcrystal can also influence the effectiveness of the latent-image center as a catalytic site for subsequent amplification during development (7,168–170). Chemical sensitizers increase the light sensitivity of the microcrystalline grains by separating photoholes from photoelectrons and from the site of latent-image formation. Furthermore, chemical sensitizers can determine the topography of latent-image formation and improve the stability of the latent-image center once formed. Typical chemical sensitizers include sulfur-containing compounds such as thiourea and sodium thiosulfate (171–178), gold-containing complexes such as gold thiocyanate and potassium tetrachloroaurate (17,18,179,180), and chemical reducing agents such as hydrogen gas and *tert*-butylamine borane [7337-45-3] (181–184). These sensitizers can be used either alone or in combination to increase sensitivity, and produce optimum photographic properties when used in trace amounts, ie, μmole sensitizer per mole Ag. At these low concentrations the sensitizers have advantageous effects on the solid-state phenomena during exposure without adversely affecting subsequent amplification processes during development.

Before the discovery of chemical sensitizing compounds, certain samples of gelatin were found to enhance the native sensitivity of silver halide grains. Subsequent investigation (171) revealed that these active gelatin samples formed silver sulfide, Ag₂S, by reaction with silver ions. Silver sulfide is generally considered to be the active chemical species resulting from sulfur sensitization treatments with either sodium thiosulfate or thiourea.

The reaction of thiosulfate with silver halide crystals to form adsorbed sulfide on the grain surfaces is activated thermally. If the reaction is allowed

to continue too long before quenching or if excessive concentrations of sodium thiosulfate are used, the emulsion grains become spontaneously developable, ie, no exposure is required to induce catalytic activity, and image discrimination is lost. The sulfiding reaction is a two-step process that involves physical adsorption followed by a thermally activated chemical reaction.



Radiotracer studies (see RADIOACTIVE TRACERS) (176) have shown that the outer-sphere sulfur atom in the thiosulfate anion is the active sulfur sensitizing agent remaining on the emulsion grains in equation 4. Actual chemical sensitization resulting in enhanced photographic sensitivity appears to require a third step in which adsorbed Ag_2S molecules aggregate to form crystalline clusters (21,185).



Based on radiotracer data, spectrophotometric analysis, and proton-monitoring experiments, the sulfiding reaction (eqs. 3 and 4) is first order in the concentration of thiosulfate and less than first order in the concentration of silver ions. This reaction proceeds with an activation energy of 84–126 kJ mol⁻¹ (20–30 kcal mol⁻¹) (176,178,186–188). On the other hand, the overall sensitizing reaction, including the third step (eq. 5) which confers enhanced photographic sensitivity to the grain, appears to be one half order (185). Compounds that strongly adsorb to silver halide surfaces, eg, tetraazaindene, interfere with the aggregation process in equation 5 and prevent optimum photographic sensitization even if the sulfiding reactions in equation 3 and 4 have gone to completion (189,190).

The mechanism by which silver sulfide enhances sensitivity of silver halide grains is related to improved efficiency during latent-image formation. It has been suggested that sulfur sensitization increases the depth of individual electron traps (139,191,192); however, sulfur may reduce the repulsive potential energy associated with the surface space charge and thereby facilitate the approach of a photoelectron to the surface for subsequent latent-image formation (193). These models share the common notion that sulfur sensitization enhances electron-trapping propensities at sites where latent image formation can occur. On the other hand, other experiments have suggested that under certain conditions sulfur sensitization can enhance hole-trapping probabilities (19,156,194–196) or can have a dual role, trapping either of the photocarriers depending on the size and/or location of the silver sulfide cluster (197–199).

Gold is often used in combination with sulfur for an additional increase in sensitivity, particularly for high intensity exposures. The role of gold in chemical sensitization is at least twofold (17,18,200,201). Gold enhances catalytic activity of the latent image center with which it is associated. In this capacity, gold reduces the total number of photochemically reduced atoms (Ag and/or Au) required for developability and therefore increases the light sensitivity of the grain. Furthermore, in association with a nucleating Ag center, gold appears to stabilize the cluster and thereby enhance the efficiency of the latent-image formation process (20,202). Gold also may assist in the sulfiding processes (eqs. 3 and 4) (202). Ag_3AuS_2 , Au_2S , and AgAuS species have been detected spectroscopically and calculations suggest that clusters of either of the first two species may have vacant electronic orbitals below the bottom edge of the AgBr conduction band and may therefore behave as traps for photoelectrons (199) and contribute to the efficiency of latent-image formation. The photographic responses for unsensitized, sulfur-sensitized, and sulfur-plus-gold-sensitized octahedral silver bromide grains with 0.7-μm edge length are shown as characteristic D-log *H* curves (Fig. 10).

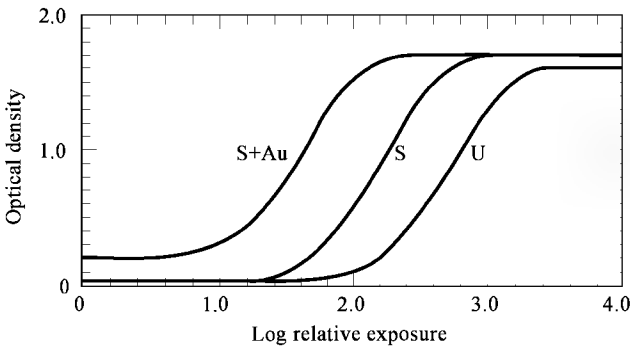


Fig. 10. Characteristic curves demonstrating the effects of chemical sensitization where curve U represents the photographic response of an unsensitized emulsion, and curves S and S + Au demonstrate the effects of sulfur and sulfur-plus-gold sensitizations. The emulsion is composed of 0.4 μm cubic AgBr microcrystalline grains coated with a gelatin binder on a film base to produce a Ag coverage of 1.25 g m⁻² of coating surface. After a 1.0 s exposure using tungsten light, the coatings are developed for five minutes in a standard black-and-white developer, fixed, washed, dried, and analyzed with a densitometer.

Treatment of emulsion grains with chemical reducing agents also enhances sensitivity. Reduction sensitization can be achieved by high pH treatments (203), low pAg treatments (204), or treatment with chemical reducing species such as H₂, stannous ions, or hydrazine [302-01-2] (qv) (205–207). Reduction sensitization contributes to enhanced sensitivity by either hole-trapping processes, electron-trapping processes, or some combination thereof, depending on the particular technique used (207–210). The hole-trapping process can provide substantial sensitivity enhancement by hole removal which reduces electron–hole recombination processes (eq. 6), and by releasing an extra electron into the conduction band which could contribute to latent-image formation. The thermal regression of atomic silver associated with the latter process is described in equation 7.



In this process a photohole, h^+ , and a reduction-sensitization center combine to produce two interstitial silver ions, Ag_i^+ , and a conduction band electron, e_{CB} . All of these products are reactants in the latent-image forming process and thus can impact favorably on photographic efficiency. However, reduction treatment of any of these types must be controlled carefully because excessive chemical reduction may produce atomic silver centers of sufficient size or activity to catalyze spontaneous development (sensitization fog). Such oversensitization can reduce sensitivity and increase the developed density in the unexposed regions of negative films. The controlled use of iodide to produce hole traps and/or as a means to establish phase boundaries which can encourage electron–hole separation has contributed significantly to the sensitivity advantages of commercial tabular grains (147,211).

Spectral Sensitization. The intrinsic absorption, and therefore the intrinsic photographic sensitivity, of silver bromide and silver iodobromide microcrystals falls off rapidly for wavelengths greater than 500 nm (see Fig. 2). In fact, silver chloride crystals have almost no sensitivity in the visible

regions of the spectrum. The need to extend silver halide sensitivity into the green (ca 500–600 nm) and red (ca 600–700 nm) regions of the visible spectrum is obvious for the production of color photographic products. Furthermore, even in black-and-white materials, extension of the photographic response beyond 500 nm is necessary for optimum effective light sensitivity. The process of expanding the wavelength sensitivity beyond the intrinsic region is called spectral sensitization.

As in chemical sensitization, spectral sensitization is usually done after precipitation but before coating, and usually is achieved by adsorbing certain organic dyes to the silver halide surfaces (47,48,212–229). Once the dye molecule is adsorbed to the crystal surface, the effects of electromagnetic radiation absorbed by the dye can be transferred to the crystal. As a result of this transfer, mobile electrons are produced in the conduction band of the silver halide grain. Once in the conduction band, the electrons are available to initiate latent-image formation.

Most mechanistic models of spectral sensitization include an electron-transfer step. Upon absorption of light by the dye molecule, an electron is promoted from the highest occupied molecular orbital (HOMO) to the lowest unoccupied molecular orbital (LUMO) of the dye. If in this excited state the electron is energetically above or close to the bottom edge of the conduction band, the electron can transfer from the dye molecule to an energy state within the conduction band of the crystal. The electron-transfer process is depicted in Figure 11. During the transfer, the dye molecule becomes oxidized and is left with a photohole trapped in the highest occupied molecular orbital. Many dyes have been discovered in which the lowest unoccupied molecular orbitals are appropriately disposed so as to promote such an electron transfer. In addition to the obvious importance of the position of the electronic energy levels, ie, HOMO and LUMO energies, structural packing, aggregation, and orientation of the adsorbed dye molecules are also factors in determining the practical photographic efficacy of spectral sensitizing dyes. In fact, such structural properties may explain why only about 30 of a possible 50,000 dyes commonly are used in the photographic industry (215).

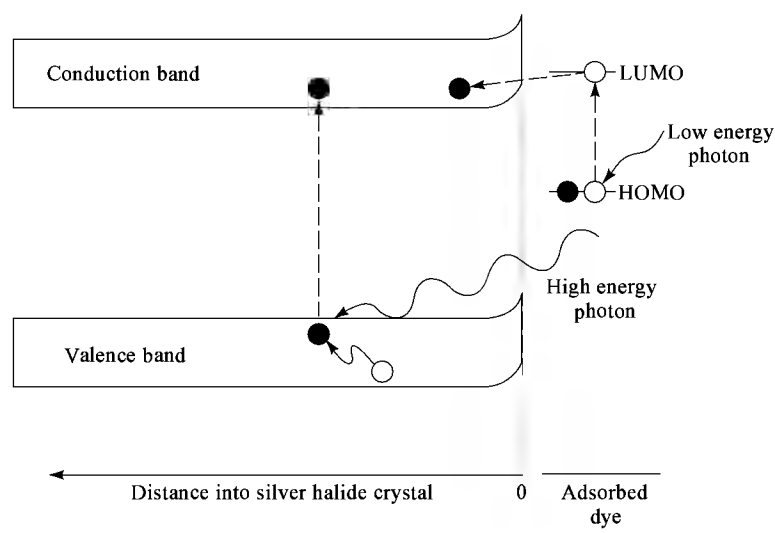
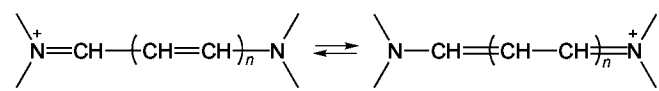


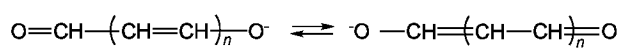
Fig. 11. Mechanism of electron transfer from an excited dye molecule to a silver halide crystal where HOMO and LUMO are highest occupied and lowest unoccupied molecular orbital, respectively.

Many spectral-sensitizing dyes can be classified according to molecular structures (228). The structural part of a dye molecule that enables the molecule to absorb visible or infrared radiation is called a chromophore. The resonance structure for three common chromophores is shown.

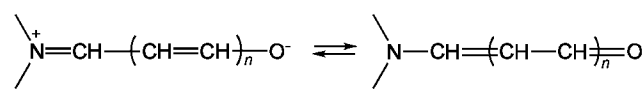
Amidinium ion system



Carboxyl ion system



Dipolar amidic system



The resonance within the chromophore corresponds to electronic motion along the length of the conjugated chromophoric chain. The terminal groups of the chromophore chain are often included in heterocyclic rings. The structure in Table 1 shows an amidinium ion system bound by two benzothiazole nuclei. If the bridge between the heterocyclic nuclei consists only of a methine group, $\text{CH}-(n=0)$, then the dye is referred to as a simple cyanine. Increasing the chain length produces carbocyanines, dicarbocyanines, tricarbocyanines, and so on for $n = 1, 2, 3 \dots$, respectively.

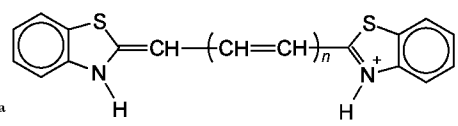


Table 1. Amidinium Ion Chromophore Series^a

<i>n</i>	Wavelength of maximum absorption, λ _{max} ^a , nm
0	423
1	557
2	650
3	758

^a Ref. 229.

The length of the conjugated chain in the chromophore and the nature of the terminal nuclei are important factors in establishing the wavelengths at

which a dye molecule absorbs incident radiation. In a vinyl series, a set of different dyes for which only n changes, the wavelength of maximum absorption increases with increasing n . Such a phenomenon can be viewed in terms of a simple quantum-mechanical model. The length of the chromophore can be viewed as the length of a one-dimensional box which confines a particle (ie, an electron). As the length of such a quantum box is increased, the energy levels come closer together (230) and thereby lesser light energies (longer wavelengths) are required to effect an electronic transition. Table 1 shows the wavelengths of maximum light absorption as a function of n in such a series of dye molecules. By structural variations, sensitivity to infrared radiation can be produced. For example, chain-stabilized pentacarbocyanine dyes have been prepared that impart photographic sensitivity to silver halide crystals at wavelengths beyond 1300 nm (231). Adsorption of the dye to a grain surface reduces the energy separation between the LUMO and HOMO. This reduction accordingly produces a bathochromic shift in the absorption maximum. There are, however, limitations for spectral sensitization in the infrared. Because the LUMO must be close to the bottom edge of the conduction band, the HOMO must move to higher energies for the sensitizing dye to absorb at longer wavelengths. These sensitizing dyes become increasingly susceptible to oxidation and, thereby, increasingly unstable (232). Accordingly, practical sensitization in the infrared is complicated by the fact that compromises between the electron-injecting ability and the thermodynamic stability of the dye must be considered. Further red shifts in the absorption and photographic sensitivity maxima occur when the adsorbed dye concentrations are high enough to promote formation of oriented dye aggregates (212). The resulting long-wavelength absorption bands are referred to as J-bands and are characterized by sharp, intense absorption peaks. The extinction and wavelength of maximum absorption for a J-band depend on the refractive index of the substrate (233) and on the crystallographic orientation of the substrate (47,48,234).

Adsorbed spectral sensitizing dyes generally produce a photographic sensitivity proportional to the number of photons absorbed. In fact, some sensitized silver halide grains promote latent image formation with the same efficiency per absorbed photon in the spectral region of the dye as in the intrinsic region of the silver halide absorption. However, the quantum efficiency is not always so high. Because of interference of adsorbed dyes with development, relative positions of the LUMO and the bottom edge of the conduction band, and or other competing kinetic pathways (216,225), certain dyes (eg, pinakryptol yellow and phenosafranine) may desensitize the photographic response, in some cases by as much as two orders of magnitude. If the LUMO of an adsorbed dye molecule is below the bottom edge of the conduction band, the dye's LUMO can desensitize the intrinsic, ie, high energy photon response by competing with the latent image centers for the photoelectrons in the conduction band. In addition to development restraint and photoelectron trapping, in some situations the photohole can cause the regression and possible destruction of the latent-image silver center. This process is favored when dyes are aggregated and can facilitate the transport of the photohole from the point of origin to the growing latent-image center. Dye photoholes have been detected and characterized by electron spin resonance (225) and reflectance spectroscopy (224,235).

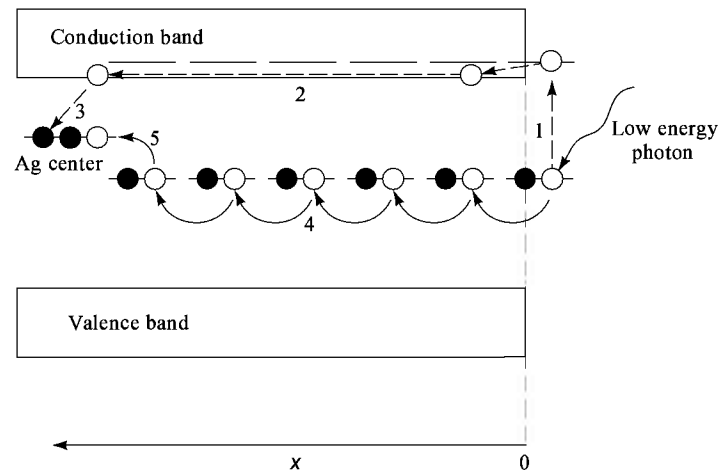
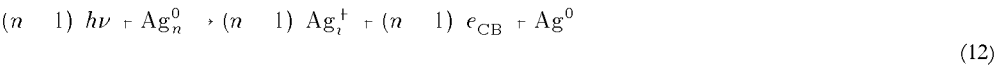


Fig. 12. A possible mechanism for the dye-induced photooxidation of a silver center. x represents the distance across a silver halide surface to which aggregated dye molecules are adsorbed. Steps 1, 4, and 5 represent the photohole (○) formation, photohole migration, and silver oxidation processes which can ultimately lead to the total regression of the silver aggregate; (●) represents an energy state occupied by an electron.

Figure 12 and equations 8–12 depict a case in which the dye molecules are aggregated on the grain surface, a condition normally considered favorable for obtaining spectral sensitization. In this case, however, the aggregated dyes can form a pathway for transporting photoholes, whether produced by an intrinsic or a spectral exposure, to a growing latent image center. Through repeated attack, steps 4 and 5 have been postulated under certain conditions to overwhelm steps 2 and 3 (Fig. 12), oxidizing the silver aggregate and eventually leading to its total regression.



If $n - 1$ photons are absorbed as in equation 8, then the complete annihilation of the n -atom silver center can occur as shown in equations 12 and 13:



where equation 8 is the result of photoexcitation of the dye and electron transfer into the conduction band, equation 9 represents hole migration from dye to dye, equation 10 describes the oxidation of the silver cluster, equation 11 represents the migration of an interstitial silver ion from the oxidized silver cluster, and equation 13 is the thermal regression (see eq. 7) resulting because the highest electronic energy state for a single silver atom is within thermal energy from the bottom edge of the conduction band.

Another common loss process results from electron–hole recombination. In this process, the photoexcited electron in the LUMO falls back into the HOMO rather than transferring into the conduction band. This inefficiency can be mitigated by using supersensitizing molecules which donate an electron to the HOMO of the excited sensitizing dye, thereby precluding electron–hole recombination. In optimally sensitized commercial products, dyes

and conditions which contribute to the many possible loss processes, eg, equations 8 through 13 and electron–hole recombination, must be avoided.

Other Emulsion Additives

In addition to chemical and spectral sensitizers, several other classes of chemical compounds are added to emulsions before coating. Additives are used to facilitate coating operations, eg, surfactants (qv) and viscosity enhancers; to reduce spontaneous development in unexposed regions, eg, tetraazaindenes and mercaptotetrazoles; and to reduce abrasion and permit high temperature processing, eg, aldehydes (qv).

For certain component compositions the viscosity and surface tension of the melted emulsion may not allow adequate emulsion spreading on the support during the coating procedures. For these situations, various surfactants that act as spreading agents are available to control the surface tension. The sulfiding reaction is a thermally activated process with an activation energy near 126 kJ mol (30 kcal mol). Therefore, quenching the reaction by cooling the emulsion from 60 to 25°C does not eliminate the reaction but rather reduces the rate by about two orders of magnitude. After long storage of the emulsion, continued sensitization may produce a catalytic activity in the silver halide grains and unwanted photographic fog upon development. This can be controlled partly by additions of such stabilizers as halide ions, acid, benzimidazoles, benzotriazoles, benzothiazolium salts (236,237), and mercaptotetrazoles (238–241). Many of these compounds adsorb to silver and complex with silver ions. Specifically as a result of these interactions, phenylmercaptotetrazole (242) restrains development (243) and enhances sensitivity (244) even in freshly coated samples. Quantum-mechanical analyses coupled with photographic data suggest that the best stabilizers not only bind with silver ions but are also poor reducing agents (245). Azaindene (246–248) compounds satisfy both of these properties and are effective stabilizers and development antifoggants. Fog control for Au-sensitized emulsions can be, in part, achieved using thiocyanate (249). Gelatin cross-linking agents (hardeners) represent another class of materials that may be added before coating. These compounds render the coated emulsion layers more resistant to abrasion during handling and improve the thermal stability of the gelatin.

The desire for reduced development times required to produce an image has necessitated the use of solutions with increased activities. Temperature increases, pH increases, and increased oxidizability are all variations directed toward shortening process times. Unfortunately, high temperature processing tends to soften and dissolve the gelatin emulsions; therefore the gelatin must be hardened before development. The enhanced thermal stability and improved mechanical durability produced by hardeners result from the formation of three-dimensional bridging of various sites within the gelatin molecules (250). Both inorganic, eg, chromium salts, and organic, eg, aldehydes, compounds have been used as hardeners. Chromium appears to complex with carbonyl groups (251), whereas many of the organic hardeners seem to cross-link between the amino groups in the gelatin molecules (252).

In most color photographic products, organic compounds such as couplers or redox dye releasers are added to the melted emulsions before coating. These compounds are essential to the development reactions that produce the dye molecules composing color images.

Coating the Emulsion

The Support. For most practical applications, the sensitized emulsions must be coated on a base or support to permit convenient handling. There are three basic classes of supports: glass, plastic, and paper. Supports are chosen on the basis of dimensional stability, low water permeability, flexibility, freedom from surface irregularities, compactness, cost, and safety. The relative importance of these requirements depends on the particular application (253). For example, dimensionally stable glass plates are one of the oldest base materials and are excellent supports for the precision photography required in astronomy, telemetry, and microelectronics. However, weight, bulkiness, and fragility make glass inappropriate and inconvenient for amateur photography. Clear plastic film supports are the most commonly used bases in modern photography. These materials are designed and selected based on safety, environmental concerns, and how they are to be used, eg, their ability to be unwound and rewound on reels and cassettes or rigidity in sheet film formats. The original plastic film supports were prepared from chemically unstable and highly flammable cellulose nitrate. Cellulose nitrate supports have been replaced by solvent-cast materials, eg, cellulose triacetate [9012-09-3], and extruded materials, eg, poly(ethylene terephthalate) [25038-59-9] (PET). These materials are not only safe but also strong and dimensionally and chemically stable. Paper supports are commonly used in products that are viewed in the reflection mode, such as color or black-and-white print materials. Before the silver halide emulsion layers are coated, some paper products are first undercoated with barium sulfate in a gelatin matrix to improve surface smoothness and visual whiteness in the highlight areas of the prints. Modern paper products are often waterproofed with impervious resin coatings of polyethylene whitened with titanium dioxide (254). Because of the water repellency of resin-coated materials, fewer chemicals are carried over from one treatment bath to another during processing. Subsequently, chemical replenishment rates can be reduced, an economical and environmental advantage (see WATERPROOFING AND WATER OIL REPELLENCY).

Coating Techniques. There are rigid constraints on the technology of coating photographic materials. The emulsion coatings must be uniform in thickness and composition and free from streaks. Compounding the difficulty is the need to produce a multilayered coating often composed of more than ten separate chemical-containing layers. Furthermore, light-sensitive materials must be coated in near or total darkness. Depending on its function, an individual, coated, dried layer is generally 1–30 µm thick (40,255–257). Most of the coating application techniques for photographic materials employ a flexible support which is transported on rollers past a coating station where the emulsion is delivered. In some systems, melted emulsion is held in troughs or trays and the emulsion is transferred to the support by bringing the moving support into contact with the surface of the liquid emulsion. As the coating rises from the trough, excess emulsion can be removed by a knife-edge or with air jets. As an alternative to trough coating, various types of coating hoppers have been developed. In the latter techniques, the liquid emulsion is delivered to the moving support by pumps or gravity flow and is spread uniformly over the support surface by variously designed hoppers (255,257). Patent literature refers to coating rates in excess of 3 m s (258). Once the emulsion has been spread over the support, the support is conveyed into a cooling chamber where the emulsion gels, then into a drying chamber where much of the water is removed from the coated gelatin. During drying, the emulsion thicknesses are reduced to ca 10⁰ % of the originally coated thicknesses.

Coating Structures. Light-sensitive photographic elements can be produced by coating an emulsion layer directly on the support; however such simple structures rarely have practical application. Generally, the support is electrostatically or chemically treated to improve the adhesion of the hydrophilic gelatin layer to the more hydrophobic support. Furthermore, in some products intermediate layers are coated between the emulsion and base to facilitate spreading. These intermediate layers often contain light-absorbing materials to prevent stray light from reflecting back into the emulsion layer during exposure. The stray light is reflected at the film–air interface on the back of the support because of the refractive index change at that surface. In the early days of photography the reflected light often produced photographs showing halos around small bright images, hence the use of light-absorbing materials is referred to as antihalation and the light-absorbing layers, antihalation layers. Several different light-absorbing materials have been used in antihalation layers, including finely divided carbon particles, dyes, and colloidal silver. The antihalation materials must be removed during processing as they can interfere with subsequent viewing. Colloidal silver dispersions are occasionally used in color products and can be removed during the silver bleaching and fixing stages. When carbon suspensions are used, the carbon is coated on the back side of the transparent film opposite the emulsion and removed by mechanical scrubbing.

In addition to intermediate layers and antihalation layers, most films also are overcoated with a gelatin layer that protects sensitive emulsion layers from the image-degrading effects of pressure and abrasion. Protective overcoat layers are generally only 1–3 µm thick. Because uv radiation is strongly absorbed by gelatin, an intervening gelatin layer would seriously reduce the effective response of the silver halide grains in materials designed for ultraviolet sensitivity (259), therefore some specialized products used in astronomy and vacuum spectroscopy do not have protective overcoat layers.

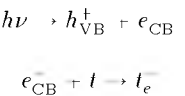
Virtually every photographic product has its own unique coating structure in order to optimize the sensitometric responses required for a particular application. To record an original scene accurately, photographic emulsions often must respond differentially to an exposure intensity range of 1 to 1000 times from the darkest to the brightest object in the scene. The exposure range over which an emulsion can respond differentially is called the latitude. Because most individual emulsions have exposure latitudes of only 1 to 100 times, two or more separate emulsions often are coated on the same support. Depending on the particular application, the two emulsions may be blended before coating and then coated in a single layer, or the emulsions may be coated in separate layers, one directly above the other (260). For color products, multiple emulsions must be coated for each of the three color-sensing regions, ie, blue, green, and red. For reflection print materials, there is often no need for long latitude response, but rather there is a need for high contrast and short latitude; therefore, emulsion blending and multiple-emulsion coating are not generally required. X-ray products have a unique coating structure with high coated levels of silver halide crystals to produce the desired x-ray sensitivity. Emulsions are coated on both sides of x-ray film, thereby providing the necessary high silver halide concentrations without severely reducing the accessibility of some emulsion grains to the processing solutions.

Exposure and Latent-Image Formation

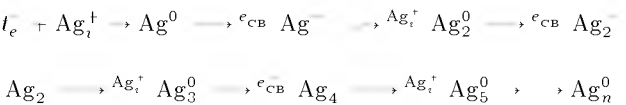
For photographic materials in which the image is produced with uv, visible, or ir radiation, optical lens systems are required. The lens system focuses the image of a scene on the emulsion layers of the photosensitive coating. The degree of magnification is, therefore, a function of the effective focal length of the optical lens system. Furthermore, the quality of the image as recorded within the coating depends not only on the optical and chemical properties of the coated light-sensitive material, but also on the lens optics. Accordingly, lens flare, astigmatism, and chromatic and spherical aberration all must be considered in designing a photographic system. Certain photographic systems do not use lenses during imaging, eg, x-ray and laser-scanning systems. In medical radiography, x-ray images are recorded by placing the photosensitive coating next to an x-ray-absorbing screen. The phosphors of the screen are electronically excited upon x-ray absorption and emit visible light during subsequent de-excitation. The original x-ray pattern is recorded as the photographic grains absorb the photons emitted from the screen.

Scene images are recorded in the emulsion layers by the action of the photoholes and photoelectrons produced in the grains by the absorption of photons. In unexposed regions of negative-working materials, there are no photoholes and no photoelectrons. Accordingly, no latent images and no catalytic silver clusters are produced and little amplification occurs in these regions during development. However, in the exposed regions, the photoelectrons react with silver ions to form clusters of silver metal. These clusters of silver function as catalytic centers for amplification during subsequent development.

Discussions and experiments regarding the theory of latent-image formation requiring analysis of events involving one, two, or three atoms generally have been based on indirect evidence (130,142,156–167). However, all of the theories begin with the same latent-image formation process, ie, photon absorption by the crystals (or the adsorbed sensitizing dyes) to produce electron–hole pairs, and all of the theoretical processes end the same, ie, with a sufficiently large cluster of localized silver atoms. Furthermore, all theories describe the formation of silver atoms as the result of electrochemical reduction of silver ions by the photoelectrons. The differences among the various theories are in the descriptions of the mechanistic steps which occur between photon absorption and final formation of a stable silver cluster, ie, latent image. In the earliest mechanistic models, the growth of the latent image was viewed to consist of an alternating sequence of coulombically driven steps beginning with the formation of an electron–hole pair and the trapping of the electron at a chemical sensitization center or a kink site yielding a localized negative charge, t_e^- :



where h_{VB}^+ represents a mobile valence band photohole and t is an electron trap. The trap is reset and able to attract a mobile interstitial silver ion, thus initiating an alternating sequence ultimately yielding a stable and developable cluster of silver atoms, Ag_n^0 .



Other models consider the steps up to and including Ag formation to be rapid in both the forward and reverse directions, thereby setting up an equilibrium condition involving h_{VB}^+ , e_{CB}^- , t_e^- , t_h^+ , Ag^0 , and Br^0 , where t_h^+ is a trapped hole and Br^0 is the result of Br[−] neutralization by t_h^+ . In establishing this equilibrium condition, recombination processes are also considered, ie, $t_e^- + h_{VB}^+ \rightarrow 0$ and $e_{CB}^- + t_h^+ \rightarrow 0$. In the nucleation and growth model from this collection of species, nucleation eventually occurs when the silver atom traps an electron and an interstitial silver ion to become a stable, Ag_2^0 nucleus. The Ag_2^0 center continues to grow by alternately trapping e_{CB}^- and Ag_i^+ . Other models suggest that the process starts when a Ag_2S center traps h_{VB}^+ , becomes positively charged, releases Ag_i^+ , and becomes AgS. Subsequent trapping of e_{CB}^- and Ag_i^+ by another Ag_2S yields Ag^0 . Latent-image centers can form either by diffusion and aggregation of the silver atoms or by collection of e_{CB}^- and Ag_i^+ at the silver atom. By whatever mechanistic process, a negative silver scale is produced (see Fig. 4) through the imagewise formation of latent-image centers and the treatment with a latent-image sensitive development solution. Negative dye-scale images can also be produced if the developer molecules, oxidized during latent-image amplification, are used in subsequent chemical reactions to form yellow, magenta, and cyan dyes.

Positive imaging can be accomplished by several different techniques. In positive imaging, the density produced by developed silver or dye molecules decreases with increasing exposure. If an original negative is on a transparent film base, a 1:1 positive copy can be produced by contact printing on a negative print material, or an optically magnified positive copy can be prepared by using an enlarger. In addition to such negative–positive approaches to producing positive images, positive or reversal imaging in the originally exposed coating is also possible. One way to generate reversal images in color films is to use specially designed processing sequences. If black-and-white development is used to develop all of the exposed grains, then a positive dye image can be produced by first chemically activating the remaining silver halide and then developing the activated grains with color developers. In the latter approach the silver halide grains are of the same negative-working type as those used in negative–positive materials. Negative-working silver halide grains are also used to produce the positive images provided by instant photography products (28). After exposure in instant photography systems, the film is passed between a pair of rollers which rupture a reagent-containing pod. The reagents include development initiating chemicals that are released from the pod and uniformly spread within the film structure. Upon reaching the silver halide-containing layers, the reagents chemically promote the imagewise diffusion of silver ions in the case of black-and-white products or dyes in color products. The diffusing species are ultimately trapped in a receiving layer for subsequent viewing (see COLOR PHOTOGRAPHY, INSTANT).

There are other reversal imaging materials, some of which make use of specially designed reversal emulsion grains (29,64,261–263). Reversal grains are conveniently divided into two classes: photobleach reversal grains and internal image grains. Photobleach grains are chemically treated to have surface catalytic activity before exposure. During exposure, the photoholes migrate to and oxidize the centers of catalytic activity. Because the oxidized centers in the exposed regions have reduced catalytic activity, development yields positive silver imaging. In internal image forming grains, exposure forms an internal latent image and deactivates the surface for subsequent nucleation treatment which produces catalytic nuclei only on surfaces of unexposed grains (29). Subsequent amplification produces a positive image in silver.

For both negative–positive and reversal materials, the optical properties of the coated silver halide layers significantly influence the quality and sensitivity of the photographic response. Once the photons enter the gelatin coating, they move on a straight path until they encounter a microcrystalline grain. At this point either the photon is absorbed by the grain and produces an electron–hole pair within the crystal, or the photon is scattered through some angle and continues its journey through the gelatin. The relative probabilities of these two events determine the vertical and lateral distribution of light within the emulsion layers (264). Accordingly, these relative probabilities also partly establish the sensitivity, sharpness, and resolving power of the photographic film. The photoholes and photoelectrons produced in the grains by the absorbed photons are mobile and free to move through the crystal; the locations at which these mobile charge carriers are trapped determines the efficiency of the photographic response. If the charge carriers are trapped so that the electrons and holes recombine, they annihilate one another and the photographic efficiency is reduced. The probability of such annihilation is markedly reduced in practical photographic films by the use of chemical sensitization and skillful crystal design.

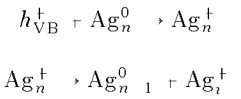
Optimized photographic grains may require fewer than 10 absorbed photons to produce a catalytic center; clusters containing only two or three silver atoms may catalyze development. The practical sensitivity or speed of such light-detection materials often is monitored in terms of the exposure required to produce a given photographic response. The exposures can be recorded on either log scales or arithmetic scales. The DIN (Germany) speeds are logarithmic, so that doubling the sensitivity of a film corresponds to a constant additive increase in the speed. The ASA (United States) and GOST (Russia) speeds are arithmetic, so that doubling the sensitivity corresponds to doubling the speed. In determining the photographic response at which the exposure is monitored, the various systems again differ. In the DIN and GOST systems, the speed corresponds to the exposure required to produce a

given optical density above fog (0.10 for DIN, 0.20 for GOST). The ASA system also considers the contrast of the response in the calculation of speed because the contrast is important in determining the quality of the recorded image. In the early 1970s an international standard (ISO) was adopted by the United States, Great Britain, and Germany (265) for black-and-white negative photographic materials. The ISO standard is an arithmetic rating which includes contrast considerations.

The developed image produced in a given photographic film is not uniquely determined by the exposure. Exposure is a measure of the total incident light energy and is therefore equal to the mathematical product of the light irradiance, *I*, and the exposure time, *t*. The developed image usually depends on the individual values of *I* and *t*; ie, a low irradiance exposure for a long time may not yield the same photographic response as a higher irradiance exposure for a short time even though the exposure, *It*, may be the same. Materials that behave in this fashion are said to suffer from reciprocity law failure. For low irradiance exposures, the time between photon absorption events is relatively long. Thus, a primary cause of low irradiance reciprocity failure may be thermal instability of a single atom of silver in a silver halide lattice. Chemical sensitization often reduces reciprocity law failure. In particular, gold is thought to stabilize isolated atoms of silver against thermal regression and thereby reduce low irradiance reciprocity law failure. Because different photographic emulsions have different reciprocity properties, the color balance of a reproduction with a multilayer coating may be altered if recommended exposure times are not used.

Special Exposure Effects

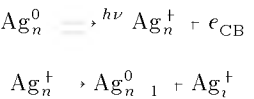
Solarization, the photobleach effect, the Herschel effect, and the Clayden effect (266) are all exposure-related phenomena that produce positive photographic images without the need for special development solutions or processing sequences. For negative-working emulsions that are conventionally processed, the developed density increases and then becomes constant with increasing exposure. For even higher exposures, the optical density and the developed silver may begin to decrease (see Fig. 4). This decrease in density is called solarization and results, at least in part, from the trapping and mobility properties of the photoholes and photoelectrons produced by high irradiance exposures. The photographic speed of positive-working films that use the solarization effect is rather low, and the phenomenon, therefore, has limited practical utility. It is generally believed that solarization results from photoholes bleaching surface latent-image centers formed earlier in the exposure. A latent-image center, Ag⁰, can be completely photobleached by repetition of the following process.



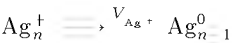
Several experiments seem to confirm the photobleach mechanism. In particular, the use of hole acceptors, eg, sodium nitrite and acetone semicarbazone, during exposure can reduce and eventually eliminate solarization. Furthermore, amplification with developing solutions that enhance the solubility of the silver halide grains also reduces or removes solarization. These results suggest that the latent images are not permanently destroyed by bleaching but merely that the topography of the centers within the grain is altered. Electron microscopy has been used to show that large, crystallized, undevelopable silver centers are found on the surface of grains in solarization regions (267), suggesting that a variety of mechanisms may operate during solarization.

The use of photoholes to bleach prefogged emulsion grains is applied commercially in the production of direct-positive images for micrographic, radiographic, and graphic arts applications. In these photobleach materials, catalytic centers are added to the grain surfaces before exposure. Subsequent exposure destroys the catalytic activity of the centers, so that positive imaging results upon conventional processing. The preexposure formation of surface catalytic centers has been accomplished by reduction sensitization and by uniform light flashes. The degree of catalytic activity imparted to the surface must be regulated carefully so that photoholes can destroy the centers during exposure. The photoelectrons, which also are produced during exposure, must be removed; otherwise electron–hole recombination or even amplification of the surface–center activity may result. The unwanted photoelectrons can be removed at the grain surface by electron trapping or desensitizing dyes, or by internal trapping sites provided by such dopants as rhodium (268) or iridium (269,270).

Ag⁰_n centers can be destroyed by the direct absorption of long-wavelength light, as well as by exposure-generated holes. If the light has enough energy, electrons can be ejected from the electronic orbitals associated with the silver centers into the manifold of electronic energy states comprising the conduction band of the silver halide microcrystal (182,271). That is,



or



where *V*_{Ag}⁺ represents a vacant silver lattice position near the Ag⁺_n center. Even though the cross section for light absorption by the small silver center, Ag⁰, is small, iteration of these steps eventually can destroy a given silver center. This phenomenon is known as the Herschel effect. With prefogged silver halide grains, the Herschel effect can produce positive imaging. In such materials the exposure of the image is carried out with red or ir radiation (266).

The Clayden effect is another exposure phenomenon that produces a positive image. This effect is produced by two sequential exposures, the first of high irradiance and the second of lower irradiance. For certain emulsion coatings, high irradiance exposures desensitize the emulsion grains, thus a second exposure produces less catalytic activity than would have been produced without the preexposure. One of the most striking examples of this desensitization was first observed in 1899 by Clayden, when he photographed lightning flashes. The short lightning flashes desensitized the grains of the negative film, and further exposure, while the camera shutter remained open, induced catalytic activity in the grains that had not been desensitized. In this manner, white images of the lightning were produced in the negative, and "black lightning" images were produced in the positive print.

Development

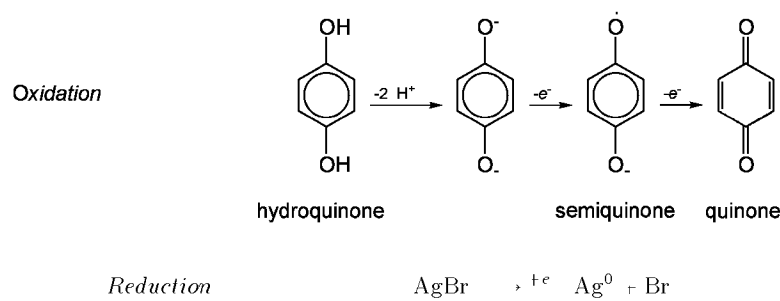
Composition of Developer Solutions. In most practical photographic materials, exposure of the silver halide grains does not produce visible images. In fact, the cluster of silver atoms composing the latent image is usually too small to be resolved with an optical or electron microscope. The growth of the small clusters of silver atoms into silver centers visible to the unaided eye is achieved with developer solutions. At least three significant classes of development components are present in most practical developer solutions: reducing agents, restrainers, and preservatives. The most important component is the chemical reducing agent.

For negative films, the electrochemical reduction properties of the reducing agents must be properly positioned to provide rapid amplification of exposed grains and much slower spontaneous amplification of unexposed grains. The ability to discriminate between exposed and unexposed grains is a well-known property of chemical reducing agents that possess the Kendall structure (272), represented by **A—(CH=CH)_{*n*}—B**, where *n* may have either zero or integral values and where A and B may be hydroxyl, amine, or substituted amino groups. Most of the useful chemical reducing agents are benzene derivatives having Kendall structures. Examples of effective compounds that have Kendall structures include hydroquinones (1,4-dihydroxybenzene), catechols (1,2-dihydroxybenzene), *p*-aminophenols, *p*-phenylenediamines, and ascorbic acid [50-81-7]. If the hydrogen of the hydroxyl group in these compounds is replaced with a hydrocarbon group, the development activity of the compound is generally destroyed. This, however, is not the case for similar hydrogen replacements on the amino groups. In fact, substitutions of the latter type enhance development activity in some cases. To be a photographically effective chemical reducing agent it is not always necessary for a compound to have the Kendall structure. In particular, phenidone

[92-43-3] developing agents and certain thiadiazoles discriminately reduce silver ions (273).

The attainment of appropriate electrochemical potentials and amplification activity in development solutions is a function not only of the chemical structure of the reducing agents but also of the other chemical components of the solution. In general, the activity of development solutions is dependent on hydrogen ion and halide ion concentrations. The pH dependence is particularly noticeable in reducing agents with hydroxyl or amino functionality. The hydrogen ion concentration often regulates the concentration of the active form of the reducing agent. Therefore, maintaining stable development activity requires pH buffering, usually on the alkaline side. During the development of silver halide grains, silver ions are converted to elemental silver and halide ions are released into the solution. The presence of released halide ions near development centers restrains amplification. Consequently, the activity of a development solution tends to decrease with increasing time of amplification. To minimize the effect of increasing halide ion concentration, development formulations usually include alkali halide salts which minimize the percentage change in the halide ion concentration during development.

Hydroquinone [123-31-9] represents a class of commercially important black-and-white chemical reducing agents (see HYDROQUINONE, RESORCINOL, AND CATECHOL). The following scheme for silver halide development with hydroquinone shows the quantitative importance of hydrogen ion and halide ion concentrations on the two half-cell reactions that describe the silver–hydroquinone redox system:



The presence of bromide ions in the development solution restrains the conversion of Ag^+ to Ag^0 by the effect of its concentration on the electrochemical overpotential for the overall redox couple. At appropriate concentrations, bromide ions restrain the rate of fog formation more than the rate of development in image areas. Such restraint selectivity usually improves picture quality, and restrainers of this type are called development antifoggants. The activity of a development solution containing hydroquinone can also be adjusted by chemical substitutions on the hydroquinone molecule. Substitution of the hydrogen atoms on the benzene nucleus with either methyl, methoxy, hydroxide, or amino groups increases the rate of development. On the other hand, the development activity is reduced by substitution with electron-withdrawing groups, eg, nitro, sulfo, cyano, carboxy, and formyl groups.

In addition to reducing agents and restrainers, preservatives are also required in practical development formulations. Unintentional oxidation of the chemical reducing agent by oxygen dissolved in the developer solution may not only alter development kinetics but may also give colored by-products that can produce unattractive stains in the developed material. It is important, therefore, to add to the development solutions chemicals that rapidly remove oxidation by-products. Sodium sulfite [7757-83-7] is an effective scavenger for such oxidation products. Sodium sulfite also decreases the rate of atmospheric oxidation of hydroquinone (274). In addition to having preservative and stain-prevention properties, sulfite has a solvent action on the grains that promotes crystal dissolution. Such solvent action influences the rate and mechanism of development.

Chemistry and Mechanism of Development. Development has been successfully viewed as an electrochemical redox reaction (275–280), for which the overall reaction can be expressed as follows:

$$n \text{ AgBr} + \text{D} \rightleftharpoons \text{D}_{\text{ox}}^{(n-m)+} + n \text{ Ag}^0 + n \text{ Br}^- + m \text{ H}^+$$

where D and $\text{D}_{\text{ox}}^{(n-m)+}$, respectively, correspond to the reduced and oxidized forms of the chemical reducing agent. The electrochemical potential of the silver half-cell is given by the Nernst equation:

$$\begin{aligned} E_{\text{Ag}} &= E_{\text{Ag}}^0 + (RT/F) \ln [a(\text{Ag}^+) / a(\text{Ag}^0)] \\ &= E_{\text{Ag}}^0 + (RT/F) \ln [K_{\text{sp}}] - (RT/F) \ln [a(\text{Ag}^0)a(\text{Br}^-)] \end{aligned} \tag{14}$$

where E_{Ag}^0 is the standard potential in volts, ie, the value of the potential at unit activity of Ag^+ and Ag^0 ; $a(\text{Ag}^+)$, $a(\text{Ag}^0)$, and $a(\text{Br}^-)$ are the activities of Ag^+ , Ag^0 , and Br^- , respectively; K_{sp} is the solubility product of silver bromide; R is the gas constant (8.314 J / (Kmol)); T is the absolute temperature; and F is the Faraday constant ($9.6485 \times 10^4 \text{ C / mol}$). The corresponding electrochemical potential for the developer half-cell reaction is given by a similar expression:

$$E_{\text{Dev}} = E_{\text{Dev}}^0 + (RT/nF) \ln [a(\text{D}_{\text{ox}}^{(n-m)+}) a(\text{H}^+)^m / a(\text{D})] \tag{15}$$

The difference, ΔE , between the two electrochemical potentials is expressed in equation 16:

$$\begin{aligned} \Delta E = E_{\text{Ag}} - E_{\text{Dev}} &= \Delta E^0 + (RT/F) \ln [a(\text{Ag}^+)] - (RT/F) \ln [a(\text{Ag}^0)] \\ &\quad - (RT/nF) \ln [a(\text{D}_{\text{ox}})] - (mRT/nF) \ln [a(\text{H}^+)] + (RT/nF) \ln [a(\text{D})] \end{aligned} \tag{16}$$

This difference is a measure of the free-energy driving force for the development reaction. If the development mechanism is treated as an electrode reaction such that the developing silver center functions as an electrode, then the electron-transfer step is first order in the concentration of D and first order in the surface area of the developing silver center (280) (Fig. 13). Phenomenologically, the rate of formation of metallic silver is given in equation 17,

$$d(\text{Ag}^0)/dt = k_f(A)[\text{D}] - k_r(A)[\text{D}_{\text{ox}}] \tag{17}$$

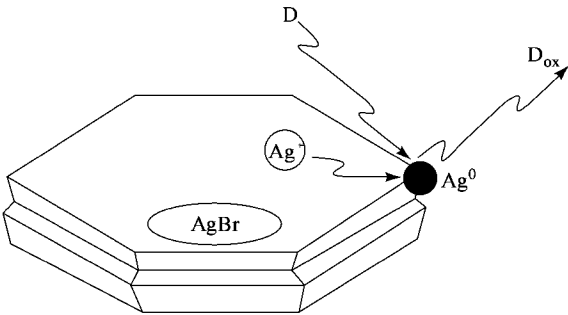


Fig. 13. A speck of developing silver on the surface of a silver halide crystal. The silver acts as an electrode for the electrochemical reduction of silver ions.

The developer molecule, D, delivers an electron to the developing speck, and in turn the electron is neutralized by an approaching silver ion which, in the process, is converted to elemental silver.

where k = potential- and temperature-dependent specific rate constants, ($\mathcal{A}$) is the surface area of the developing silver center, $[D]$ is the concentration of the reduced form of the chemical reducing agent, and $[D_{ox}]$ is the oxidized developer concentration. By using the Butler-Volmer equation, one can obtain an expression for k in terms of the overpotential, ΔE , and the temperature (281). Specifically,

$$k_f = k_f^0 \exp\{\alpha n F \Delta E / RT\}$$

(18)

and

$$k_r = k_r^0 \exp\{(\alpha - 1) n F \Delta E / RT\}$$

(19)

where k_f^0 and k_r^0 are the specific rate constants when $\Delta E = 0$ and α is the potential symmetry factor. Combining equations 16–19 generates an expression that predicts the dependence of development rates on halide ion, hydroxide ion, and reducing agent concentrations. Most analyses using the electrochemical model use additional semiempirical parameters. The temperature dependence of development is not deduced readily from the above equations. In most cases, empirical analyses have revealed that the development rate increases with increasing temperature. The observed temperature dependence and the corresponding activation energies suggest that the processes are diffusion controlled, particularly for rapid developers (282). Because processing rates are enhanced at high temperature, many modern systems have been designed for high temperature processing. The use of high temperatures has necessitated specially hardened coatings to avoid remelting of the gelatin during processing.

Initiation of development also depends on the diffusion rates of the developer components through the gelatin to the silver halide grain surfaces where the catalytic silver centers are located. It has been suggested that such diffusion processes are rate limiting (283,284). Once a chemical reducing agent has diffused to a grain surface, nucleation and growth of an elemental silver phase can be initiated in unexposed regions. In the exposed regions, the Ag^0 nuclei, ie, latent-image centers, that were photochemically produced continue their growth via chemical amplification. If the activity of the silver in the developing center, $a(Ag^0)$, is assumed to be unity, the calculated ΔE values for the initiation of development frequently are greater than those required to sustain development, once started. The silver specks constituting the latent image appear to have activities greater than unity, in accord with the Gibbs-Thomson equation (285,286). Such effects may explain the requirement for larger ΔE values during nucleation and early growth. Furthermore, the ability of a developer to discriminate between exposed and unexposed grains has been attributed to this inverse relationship between silver center activity, $a(Ag^0)$, and silver center size (286–288); ie, as the size of the silver center increases, ΔE , the specific rate constant, and the rate of development increase (see eqs. 16–19). As development proceeds, oxidation by-products, eg, D_{ox} and Br^- , build up. These components decrease the overpotential, ΔE , and thereby reduce the rate of development. Thus, maintaining high development rates requires diffusion of oxidation products away from development centers. Under development conditions for which diffusion processes become important, effective agitation of the development solution is required. In general, agitation improves the uniformity of development and enhances the overall development rate. Solutions can be agitated by bubbling nonreactive gases, eg, nitrogen (qv), through the developer solution by mechanical paddles that sweep back and forth near the coated surfaces, or by jets that force streams of developer across the coated surfaces.

Source of Silver Ions During Development. For rapid electrochemical reduction of silver ions throughout development, high silver ion activities must be maintained (see eq. 16). The silver ions may come from the development solution, directly from the silver halide crystals, or indirectly from the crystals after crystal dissolution. If amplification is achieved with silver ions from the development solution, the amplification process is called physical development. Most physical development solutions are thermodynamically unstable because of the simultaneous presence of silver ions and a reducing agent. In most practical photographic materials, the silver ions in the silver halide lattice serve as the source of silver ions during amplification. If the silver ions migrate directly to the developing center via solid-state diffusion, without venturing into the solution, the process is called chemical or direct development. On the other hand, if the silver ions from the silver halide pass through the solution on their journey to development centers, the process is referred to as solution physical development. Some studies suggest that silver halide dissolution before chemical reduction of the silver ions is particularly important during the early stages of development (289–291).

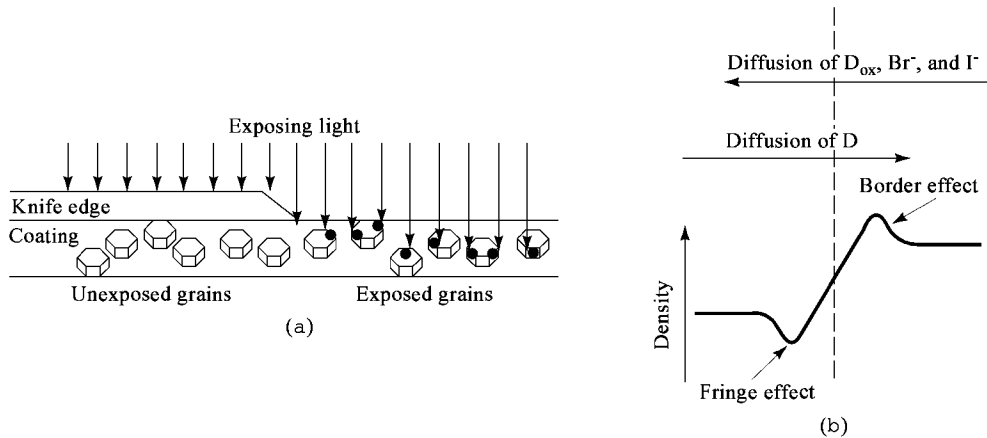


Fig. 14. Edge effects produced during development: (a) schematic and (b) a microdensitometric trace across a developed edge resulting from a knife-edge exposure which reveals the effects of diffusion processes during development. See text.

Adjacency Effects During Development. The presence of photographic grains at one point in a coating can influence significantly the behavior and response of neighboring grains. During exposure, such phenomena can occur because of light scattering. The path of a photon entering a film coating is not necessarily confined to a straight line; collisions with coated silver halide grains may scatter the photon several times before it is absorbed by a grain. The lateral photon displacement, which may result from scattering events, reduces the visual sharpness of the recorded image. Interactions among the grains also occur during development. Because of the latter interactions the optical density produced upon development is a function not only of exposure but also of development characteristics. If two adjacent areas of a coated photographic film are given exposures differing significantly in magnitude, then development phenomena in one area may influence the course of development in the adjacent area. For a negative-working film, the area receiving the greatest exposure develops more rapidly, and accordingly a relatively high concentration of development by-products, eg, oxidized developer and bromide ions, is produced. As these by-products migrate into the adjacent areas that received lower exposures, development is relatively retarded and a light line or fringe is created. Similarly, the relatively high concentration of fresh developer in low exposure areas diffuses into high exposure regions and produces a dark line or border as a result of enhanced amplification rates. This edge effect phenomenon is diagrammed in Figure 14 for a knife-edge exposure. The parallel light and dark lines generated in a coating by knife-edge exposures are often called Mackie lines. Intergrain diffusion of development components and by-products is also monitored by line exposures of various widths, usually in a range from micrometers to millimeters. Line exposures are

produced by either visible or x-ray radiation. Because x-rays are not scattered by silver halide grains, the adjacency effects observed for low (ca 30×10^{-16} J) energy x-ray line exposures are a result solely of development phenomena. On the other hand, the observed adjacency effects resulting from light-line exposure are a consequence of the superimposed contributions from light-scattering and developer component diffusion. Figure 15 shows microdensitometry tracings for x-ray line exposures of 10- μm and 1000- μm line widths. The enhanced density in the 10- μm line relative to the 1000- μm line is an example of the Eberhard effect, which was discovered in 1912 upon photographing bright astronomical bodies of various sizes (292). The densitometric traces of edge effects and Eberhard effects are mechanistically useful in establishing the role of diffusion during development.

For some developer formulations, partly oxidized or partly exhausted developer may be more active than fresh developer, eg, formaldehyde–hydroquinone developers. For these situations, the photographic adjacency effects are opposite to those observed in the Eberhard and Mackie line effects. In particular, grains in the neighborhood of developing grains develop more rapidly than they would otherwise. This phenomenon of development by-products sensitizing and accelerating the development of adjacent grains (infectious development) can lead to image spread and high contrast (293,294). Compounds such as formaldehyde (qv) and hydrazine, when added to hydroquinone developer solutions, appear to promote the electrochemical reduction of the fully oxidized developer (quinone) to the semiquinone form, which is then free to migrate and act as a development initiator. Formaldehyde seems to promote semiquinone formation by buffering sulfite concentrations; hydrazines appear to promote semiquinone formation directly by electrochemically reducing quinone to semiquinone.

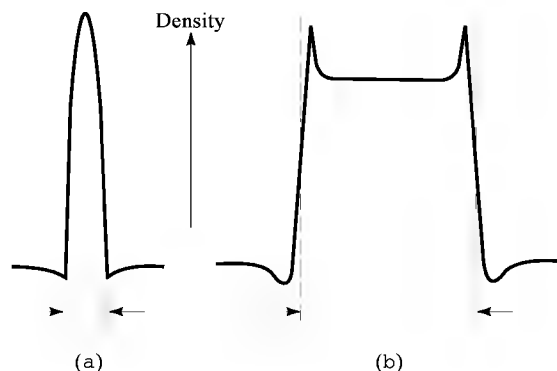


Fig. 15. Microdensitometer tracings of x-ray line exposures with line widths of (a) 10 and (b) 1000 μm .

Development-induced adjacency effects frequently enhance edge contrast, which translates into desirable sharp and crisp photographs. Such chemical sharpness often compensates for the sharpness losses produced by light-scattering during exposure (34). The apparent sharpness of developed images is also influenced by certain gelatin–developer interactions, eg, the oxidized form of the developer pyrogallol [87-66-1] (1,2,3-trihydroxybenzene) acts like a hardening or tanning agent and shrinks the gelatin. Therefore, in regions of high exposure a considerable concentration of oxidized developer is generated and a corresponding differential shrinkage of gelatin occurs upon drying. The use of tanning developers may offer yet another approach to visual sharpness control.

Special Development Methods. In addition to the negative–positive and reversal process sequences, several photographic materials have been produced based on special development techniques including activator processing, photodevelopment, thermal development, monobath processes, and silver diffusion transfer. To reduce the chemical content and thereby improve the stability and oxidation resistance of development solutions, materials that use an activator process have been developed. The light-sensitive materials used in activator processes have developing agents incorporated in the coated layer during manufacture. Incorporation of developing agents during coating without the production of severe photographic fog is possible because the coating pH is kept low enough to inactivate the developer. After exposure, the coating is developed by immersing it in an alkaline activator solution. The initiation of development in the activator solution is more rapid than in conventional processes because the developer molecules need not diffuse into the light-sensitive layers from the processing solution. In spite of the low activity of the coated developer, some unintentional reduction sensitization may occur, which produces unwanted fog. Therefore, coating the developer in a separate layer usually is preferred. Because of simplicity, rapid access, and solution stability, incorporated developer papers have been used for office copying applications.

Continued exposure to light can convert enough silver halide photolytically to metallic silver to produce a satisfactory image. This approach to photographic imaging is used for specialized applications (295). For certain emulsion grains, photolytic amplification under low irradiance produces a visual density, provided the photolytic amplification was preceded by a high irradiance exposure. In regions of the light-sensitive element that received high irradiance pre-exposures, photodevelopment produces a higher covering power (ratio of optical density to developed silver) compared with background areas. Electron micrograph examinations of the image and background areas reveal that the covering power effects result from a higher dispersity of developed specks per grain in the image areas. The discrimination between the image and the background of photodeveloped films may be a consequence of color differences resulting from the diverse shapes of the developed silver particles. The images produced by photodevelopment generally are not permanent; however, specially formulated bathing treatments have been devised to stabilize the photodeveloped images. Photodevelopment materials provide a rapid, solution-free approach for recording data from the output of oscilloscopes and other scientific instrumentation.

Heat or thermal development is another example of amplification that does not require the use of aqueous development solutions. Photothermographic materials are developed by heating the coatings after image exposure (296). In these materials, all of the required chemicals are incorporated into the coated layers during manufacturing. There are several types of heat-processed materials that are silver based; some are quite similar to conventional silver halide materials. The light-sensitive components are silver halide microcrystals and amplification is effected by the incorporation of a reducing agent that is thermally activated. Some thermally processed materials are amplified at least in part by physical development. In addition to coated silver halide and reducing agent, these materials also have an incorporated source of silver ions, usually in the form of organic silver salts. Another thermally processed material is based on silver oxalate. The oxalate material does not require an incorporated reducing agent; silver oxalate thermally decomposes to form metallic silver in a reaction catalyzed by photoproducts formed during image exposures (297,298).

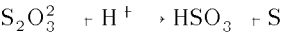
In most conventional processing, development is completed before the start of fixation, which removes the undeveloped residual silver halide. However, in monobath systems, development and fixation occur simultaneously and not sequentially. Monobath processing is a time-saving procedure in which large quantities of a silver halide solvent, eg, thiosulfate, are added to a development solution. The solvent forms a water-soluble silver complex which transfers silver from the coating to the solution. The developer component of a monobath solution must be very active to ensure that latent-image sensing and initial amplification occur before crystal dissolution. Balancing the relative concentrations of developing agent and solvent and controlling the temperature can give the desired sensitometry. Because of the pronounced adjacency effects that occur in monobath processes, images with good sharpness and high resolving power have been produced. A further extension of the monobath process allows the formation of positive images. The positive image is formed by laminating a monobath-processed coating to a receiving layer containing silver nuclei coated on a separate support. The solubilized silver ions from the original coating are thus permitted to diffuse to and grow via physical development on the nuclei in the receiving layer. Because the originally developed silver remains immobilized in the original coating, subsequent separation of the laminated layer yields not only a positive silver image in the receiving layer but also a negative image in the original layer. This process of silver diffusion transfer has been used for document copying, aerial photography, instant photography, and in graphic arts for the preparation of prepress materials (299–301).

Stop-Bath Treatment and Fixation

Once development has progressed to some desired extent, amplification usually is quenched by a rapid decrease in pH. In conventional photographic films, the pH is decreased by mechanically or manually transferring the coatings from the alkaline developer solution to an acidic stop bath. Quenching usually can be done by a brief treatment with a stop bath composed of a dilute solution of a weak acid, eg, 0.5% acetic acid. For both the silver diffusion-transfer films and the dye-transfer films used in instant photography, development is initiated by spreading a viscous alkaline reagent between an emulsion layer and an image-receiving layer (26–30). In the early versions of these products, development was stopped by a peel-apart step in which the processing chemicals and developing grains were physically separated from the image-receiving layer; more recent integral films rely on the timed release of an acid to quench development. In instant photography, the final image, whether dye or silver, is formed by diffusion. In this process unwanted chemicals that could potentially interfere with viewing and image permanence are left behind. On the other hand, in conventional photography the image remains where it was originally formed. Thus, the image and the by-product chemicals are in the same layer after development has been quenched. In the latter films, post-development procedures are required to remove the unwanted chemical by-products. To remove undeveloped silver halide, development and quenching are followed by fixation. Most fixing baths are composed of thiosulfate ions formed by dissolving the corresponding sodium or ammonium salt in water. The thiosulfate ions convert the remaining silver halide to water-soluble complexes such as argentodithiosulfate, $\text{Ag}(\text{S}_2\text{O}_3)^{3-}$, and argentotrithiosulfate, $\text{Ag}(\text{S}_2\text{O}_3)^{5-}$. Once bound in these complexes, the silver can be washed readily from the gelatin coating. During the dissolution of the residual silver halide grains, silver ions break away from the lattice of the crystal surface and pass into the solution in the form of complexed ions. In the absence of thiosulfate, the concentration of silver ions continues to increase; eventually the solubility product of the silver halide is reached, and accordingly dissolution ceases. However, by sequestering silver ions and forming various argentothiosulfate complexes, the use of thiosulfate-fixing solutions permits continued dissolution and eventual removal of all of the undeveloped silver halide. Efficient fixing agents must have large stability constants (302); ie, they must be ligands having a sufficiently high capacity for complexing with silver ions. The stability constants at 25°C for $\text{Ag}(\text{SO}_3)^{5-}$, $\text{Ag}(\text{NH}_3)^{+2}$, $\text{Ag}(\text{SCN})^{2-}$, $\text{Ag}(\text{S}_2\text{O}_3)^{5-}$, and $\text{Ag}(\text{CN})^{2-}$ are ca 8×10^8 , 1.6×10^7 , 2×10^{10} , 8×10^{13} , and 8×10^{21} , respectively. Cyanide is an effective complexing agent because of the large stability constant, however health and safety factors prohibit its use in most situations. Thiosulfate is nontoxic, relatively inexpensive, and nonreactive with the gelatin or developed silver, in addition to having a satisfactory stability constant. For the fixation of silver chloride, silver bromide, and silver bromochloride, thiosulfate is effective. However, for the relatively insoluble silver iodide with a solubility product of ca 10^{-16} , dissolution rates with thiosulfate solutions are significantly diminished.

In black-and-white photography, fixation generally is conducted under acidic conditions. The acid environments assist in arresting continued development and in preventing stain associated with the oxidation of developer molecules carried over from the development solution. Furthermore, weak acidity is required for the hardening of gelatin by potassium alum and chrome alum, which are components that can be included in acidic fixing baths. The hardening provided by alum helps prevent physical damage, eg, scratches on the coating surfaces, during handling of the wet gelatin materials in the course of various post-development processes. The acidity can be maintained if, in addition to the salt of a strong base (eg, the sodium ion from sodium thiosulfate and sodium sulfite), a weak acid (eg, acetic acid) is present to provide a buffered solution. In alkaline thiosulfate fixing baths, alum forms an undesirable white sludge of aluminum compounds.

The primary component of a fixation bath, thiosulfate, tends to decompose in acidic environments according to the following reaction:



Although this decomposition is not substantial at pH values greater than 4, approaches have been developed to protect the thiosulfate should more acidic conditions occur. The decomposition rate and the accompanying reduction in fixation efficacy can be retarded significantly by additions of bisulfite. Bisulfite ions disproportionate in the presence of thiosulfate (303). This disproportionation therefore balances the decomposition of thiosulfate and extends the useful lifetime of the bath. Besides silver-complex-forming agents, hardeners, bisulfite, and buffers, some fixing baths also contain boric acid, which increases the fixing bath lifetime by extending the pH range of effective hardening to higher pH values. This reduces sensitivity to the adverse effects of alkaline carryover from the development solution. Boric acid also decreases the propensity for alum hardeners to initiate sludge formation.

In climates where temperatures and humidity are high, extreme hardening of the gelatin generally is necessary to avoid surface abrasion of the processed coatings. A solution of potassium chrome alum (potassium chromium sulfate) can be used in these situations as an intermediate treatment between development and fixation. In the presence of organic acids, eg, acetic acid, chrome alum solutions do not attain their optimum hardening capacity. The required acidity for optimum hardening by chrome alum baths is achieved with sulfuric acid, which does not produce a buffered solution. Therefore, bath lifetimes are limited and frequent acid additions are required. For environmental reasons, the chromium hardeners are no longer preferred and hardening through the use of aluminum ions in fixing baths has become increasingly popular.

The rate of silver halide dissolution and of the ultimate total removal of undeveloped silver from a coating depends on the nature of the coated emulsion, the coating format, the thiosulfate ion concentration, the nature of the thiosulfate's cation, and the temperature and degree of agitation (304). Ammonium thiosulfate [7783-18-8] solutions remove silver halide faster than equivalent concentrations of sodium thiosulfate solutions; for this reason the ammonium compound is often preferred for the fixation of camera-speed emulsions, which are often composed of grains with relatively insoluble iodide-containing phases. Temperature increases enhance the rate of silver removal by increasing the rate of diffusion. Increased temperatures also increase the chemical reaction rates of crystal dissolution and complex ion formation. Dissolution rates can be increased up to an optimum value by increasing the concentration of the complexing agent. Increasing the complexing agent concentration beyond the optimum generally results in reduced fixing rates. For most practical analyses, the rate of fixation is monitored in terms of the clearing time, which is defined as the time required for the last visible opacity to disappear. Generally, the clearing time increases with increasing coated layer thicknesses and increasing sizes of the silver halide grains. Clearing times are inversely related to fixation rates. In conventional color films the developed silver is not used for the final image. In these films the developed silver is bleached, ie, oxidized with an agent such as ferric ion, before fixation and nearly all of the originally coated silver can be cleared and recovered from the coating.

Washing

At various stages during the post-development process, the coatings are rinsed. One of the most important reasons for rinsing is to reduce chemical carryover from one treatment bath to another, which increases solution lifetime. The most important washing occurs at the end of the processing, just before drying. The purpose of this washing is to eliminate all soluble compounds from the coated gelatin layers. Efficient removal of certain compounds is essential for good keeping properties and image permanence. Thiosulfate and argentothiosulfate complexes have particularly damaging effects on the keeping properties of the final print or film. Residual thiosulfate eventually reacts with the silver image in black-and-white products to produce silver sulfide and thereby confer a yellowish brown appearance or tone to the image. Inadequately removed silver thiosulfate complexes gradually decompose and are converted to silver sulfide with an accompanying yellow stain. For photographic materials coated on water-impermeable bases, such as glass plates and film, efficient washing can remove unwanted soluble residues almost completely. However, it is difficult to completely desorb and remove thiosulfate and argentothiosulfate complexes from the paper supports and baryta layers of reflection-print paper products. Resin-coated paper bases have made possible the manufacture of print materials having washing properties similar to those of films and plates.

The rate of washing primarily depends on the rates of diffusion of the water-soluble species in gelatin. Because the process is diffusion controlled, washing is accelerated by temperature increases and by maintaining a fresh layer of water at the film surface. Therefore, good agitation is necessary and jet washing is particularly effective. Reticulation of the coated gelatin occurs when the temperature of the wash water is excessive. Because the wrinkling of reticulation remains even after drying, this effect establishes an upper limit to wash temperatures. Washing rates are also affected by pH of the wash bath, hardening that may have occurred in previous steps, and the presence of certain salts in the wash bath (305). At pHs below the isoelectric point of gelatin, the gelatin assumes an overall effective positive charge, which appears to reduce coulombically the rate of removal of various negative species, including thiosulfate and the argentothiosulfate complexes. Thiosulfate is desorbed more rapidly from gelatin that has been previously hardened with potassium

chrome alum than from gelatin hardened with potassium alum. The rate of thiosulfate elimination can also be enhanced by the use of wash baths containing sodium chloride. However, a brief fresh water rinse is required to remove residual chloride, which might otherwise have adverse effects on the image silver in black-and-white films.

Stabilization

Stabilization is an alternative to fixation for the production of a permanent although probably not archival image. In stabilization, the undeveloped silver halide is not removed but rather is converted to a compound that is relatively insensitive to light and stable to heat, humidity, and atmospheric gases. As in fixation, stabilization is achieved by the use of complexing agents to transform silver halide into silver complexes. Because the silver complexes and the unreacted complex agents are retained in the coated gelatin layer, the stability of these compounds is critical to the utility of the stabilization process. Stabilization is used where rapid access is needed, eg, news film and oscillograph tracing. Thus the complexing agents should react rapidly and completely with the silver halide contained in the gelatin layers. Furthermore, the agents cannot be toxic and should not soften the gelatin, bleach the image silver, or form colored complexes with silver.

Table 2. Worldwide Producers of Sensitized Materials

Company	Country
Eastman Kodak Co.	U.S.
E.I. du Pont de Nemours & Co., Inc.	U.S.
Minnesota Mining and Manufacturing Co.	U.S.
Polaroid Corp.	U.S.
Fuji Photo Film Co.	Japan
Konishiroku Photo Industry, Inc.	Japan
Agfa-Gevaert, Inc. ^a	Germany Belgium
Ilford Ltd. ^b	U.K.
Ciba-Geigy Photochemie Ltd.	Switzerland
Anitec ^b	U.S.

^a Subsidiary of Bayer Chemical (Germany).
^b Subsidiary of International Paper (U.S.).

Stabilizing agents can be classified into two categories: those agents that form water-soluble silver complexes and those that form insoluble complexes. The former category includes thiocyanate, thiosulfate, and thioureas. The latter category includes mercapto compounds, ie, RSH, where R represents any aliphatic, aromatic, or heterocyclic group (306) (see THIOLS). The rapid-access print process generally uses specially prepared papers containing incorporated developing agents. The papers are processed by an activator–stabilizer sequence requiring only a few seconds of machine process time, and the final drying of the prints is in room air (306–308).

Economic Aspects

The conventional photographic industry continues to prosper. The Gross National Photo-Product (GNPP), defined as the shipment of photographic equipment and supplies by American manufacturers minus exports plus imports, is a measure of the photographic goods consumed in the United States (309). The GNPP increased from \$4.1 × 10⁹ in 1971 to \$20.9 × 10⁹ in 1991, which corresponds to an average annual increase of ca 8.5° 0, within one-tenth of one percent of the average annual gross national product growth rate over the same period. In 1991 in the United States, the total number of pictures taken was greater than 21 × 10⁹ and approximately 2.1 × 10⁶ kg (66 × 10⁵ oz.) of silver were consumed by the photographic industry. Worldwide the number of pictures taken was about 56 × 10⁹. The United States makes up ca 41.6° 0 of the international photomarket. Japan, with 27.8° 0 of the worldwide market, and Western Europe (excluding Germany), with 10.9° 0, consume 1.94 and 1.91 × 10⁶ kg (61 and 60 × 10⁶ oz.) silver, respectively. Germany (6.6° 0), Eastern Europe (10.3° 0), and others, including Africa and Asia (2.8° 0) make up the remaining countries in the international market. Table 2 gives a partial list of worldwide producers of sensitized photographic materials.

Environmental Aspects of Processing

The quality of industrial effluents discharged into public sewage systems is specified by regulatory agencies. For certain quality parameters, photographic processing effluents fall within the required range without special treatment. For example, the pH of processing wastes generally is 6.5–9.0 and thus within the usual sewer code range of 5.5–9.5. Lubrication oils and greases are not present in photographic wastes, and the concentrations of suspended solids are too low (generally <20 mg L) to be significant. Occasionally, specific processing baths must be maintained in a range of 38–51.5°C. However, when the total effluent is mixed, the temperatures are usually below 32°C, and thermal effects are generally of little consequence. Despite these qualities of photographic waste, processing effluents that ultimately are discharged into streams almost invariably require special treatment to meet stream standards. These treatments normally include silver removal, settling, biochemical degradation, aeration, and finally chlorination. Settling removes the solid wastes, biochemical degradation and aeration reduce the biochemical oxygen demand (BOD) of the waste, and chlorination destroys any pathogenic organisms remaining after treatment.

The concentrations of chemicals found in photographic processing effluents generally are not toxic to the bacteria and other treatment plant microorganisms necessary for biological degradation (310). However, hexavalent chromium used in dichromate bleaches as an oxidizing agent to remove developed silver from color coatings can be toxic to bacteria. It must, therefore, be electrochemically reduced before discharge. When the waste bleach is mixed with waste fixation and development chemicals, electrochemical reduction of hexavalent chromium generally occurs, and the adverse bactericidal properties are eliminated. Some processing components that may have undesirable environmental effects can be recycled during photographic processing and are not discharged into waste-disposal systems. Wash water, developer, and fixing and bleach-fix solutions are often regenerated, recycled, and reused. Silver can be electrochemically removed from thiosulfate fixing solutions which can then be chemically readjusted and reused (311). Many color-processing sequences have used ferricyanide as a silver bleaching agent, however, the ferricyanides hydrolyzed in the presence of oxygen and sunlight produce iron hydroxide and soluble cyanides which are toxic to fish at levels as low as 1.0 mg L (312). Because of adverse environmental impact, some of these solutions have been replaced by bleaches containing iron(III) ethylenediaminetetraacetic acid. For economic and environmental reasons, there has been an elevated emphasis on recycling, reuse, and recovery since the 1970s (313) (see RECYCLING). Furthermore, the use of thinner and harder coated layers in photographic products has reduced the carryover from tank to tank during processing. Accordingly, solution replenishment rates and the total effluent volumes have been substantially reduced (314). Effluent codes for many areas regulate not only the concentrations of certain chemicals but also the BOD and chemical oxygen demand (COD) in effluents. Photofinishing businesses occasionally install aerated lagoons capable of reducing the oxygen demand of their waste before discharge into community sewer systems. Such lagoons must be carefully engineered to guarantee that subsurface soil and streams do not become chemically contaminated.

Compounds containing such metals as copper, barium, lead, molybdenum, and nickel generally are not used in processing solutions. However, trace quantities of certain metal dopants occasionally are used to impart desired solid-state and photographic properties to emulsion grains. Because of its

toxicity to aquatic life and microorganisms, the metal of primary ecological concern in the photographic industry is silver. Silver that is not salvaged during recovery operations may be present in the waste from photographic processing. The discharged silver usually is bound in thiosulfate complexes, which are not detrimental to the essential microorganisms in sewage treatment plants. The silver thiosulfate complexes can be converted to insoluble silver sulfide and removed as a solid sludge. Optimization of silver recovery is not only an important conservation measure but is also an economic benefit to photographic processing businesses. Various techniques for silver recovery are in common use including electrolysis, metallic replacement, ion exchange (qv), and silver precipitation. Although relatively costly, electrolytic recovery does produce high purity silver plated on an electrode. Metallic replacement can be accomplished by passing the waste fixation solution through a filter cartridge containing spun steel (steel wool), iron mesh, or iron-coated fiber glass. Because iron is more electropositive than silver, it is ionized and dissolved into the solution; silver ions are reduced and precipitate as solid metallic silver sludge. Most reclaimed silver is sold to industries that manufacture photographic goods. Not all photographic processing facilities take advantage of this recycling process. Only about 50% of the silver is recovered. This percentage is expected to continue to increase because of the market value of silver and costs of noncompliance with governmental regulation.

The Silver Image

Stability. Image permanence and archival quality of recorded information (315) is of increasing interest. Recording information using silver, a relatively inert noble metal, is one of the most effective archival approaches available. However, after the silver image of a black-and-white processed material is produced, it is usually subjected to a variety of environmental conditions during subsequent handling and storage. In addition to the effects of residual thiosulfate and silver thiosulfate, the humidity, temperature, and chemistry of the environment can also adversely affect image permanence. The elemental silver composing the image is not completely stable and may be oxidized by the atmosphere to silver sulfide or silver oxide. Because such oxidation is initiated on the surface of the developed silver image, image degradation by air oxidation is of particular concern in photographic materials containing grains with high surface-to-volume ratios, especially if keeping properties are desired. For example, some microfilm images for which image permanence is often particularly important are susceptible to spontaneous image degradation upon normal storage (316). The gelatin binder also can influence image-keeping qualities. Because of the shielding effects afforded by gelatin vehicles, most modern photographic materials are considerably more resistant to air oxidation than were the early daguerreotype and calotype images. However, various degrees of image fading can still occur even in large grains with gelatin protection. If the silver image is exposed to oxidizing fumes from such sources as automobile exhaust, freshly painted rooms, nitric oxides, and peroxides, then the possibility of image fading is enhanced.

During aging, some microfilms develop red or yellow spots 10–150 μm in diameter. These blemishes have been associated with the disappearance of the silver filaments formerly composing the image and the ensuing formation of small spherical particles of silver. The mechanism of silver transport is likely to involve the oxidation of silver metal to silver ions, which migrate through the gelatin matrix and ultimately regather to form isolated colloidal particles of silver. The coloration of the blemish is at least in part a result of the fact that light scattering by a colloidal particle of a given size is wavelength dependent. The coloration may also be affected by a thin silver sulfide shell, which has been observed to reside on the silver in the blemish areas. Occasionally, Liesegang ring structures have been found to be associated with such blemish spots (316). To improve silver image resistance to oxidation, photographic coatings can be treated with solutions such as gold chloride-plus-thiourea solutions or with iodide either during or after the process. Adsorbed iodide appears to stabilize the silver image by reducing the surface energy of the silver filaments (317). Selenium toning of microfilm products has been shown to enhance image stability (318) (see SELENIUM AND SELENIUM COMPOUNDS).

Image Tone. The morphology and size of developed silver depends on factors relating to the emulsion grain as well as the method of development and composition of the developer. Developed silver is often in the form of long, narrow, cylindrical filaments. The filaments are commonly 15–25 nm in diameter but may be as large as 50–80 nm with slow-acting developers such as ascorbic acid (280). Development in environments that promote solution physical development tends to enhance the filament thickness. The filament dispersity, ie, the number of filaments per grain, depends in part on the grain size. In general, increases in the silver halide grain size translate into increased filament dispersity. For sufficiently small grains, development can lead to only a single filament per grain (319). In addition to grain size increases, exposure irradiance increases also have been observed to increase filament dispersity. For large grains exposed with high irradiance, a mass of developed filaments can be produced. Under an electron microscope, such filament masses resemble steel wool. In some cases the morphology of the silver halide grain itself, as well as adsorbed chemical compounds and trace impurities, have influenced the morphology of the developed silver (320–323). At sufficiently low concentrations of developed silver within the gelatin matrix of the coating, the spectrum of the light absorption is dependent on filament size and morphology. However, when the volume concentration of silver exceeds 3%, the coatings become visually neutral and the effect of particle size lessens (324).

Once the silver is developed, changes in image tone are sometimes desired. Treatment with any of a wide variety of toning baths can alter image tone. Two types of toning baths are used: one type involves the conversion of the image silver to silver sulfide, and the other involves the substitution of silver in the image with yet another metal. Metal toning makes possible a wide range of image colors. For example, blue and red tones can be achieved using gold substitutions, yellow and green tones are obtained using vanadium substitution, and sulfide and selenide toning can produce brown tones. Sulfide toning can be accomplished directly by treatment with a heated bath containing sodium thiosulfate and alum. An alternative method of sulfiding involves two steps. The silver image is first bleached to form silver bromide, which is then converted into silver sulfide. The bleaching can be accomplished by a treatment with a solution containing potassium ferricyanide and potassium bromide. Conversion can then be effected by imbibition either in a dilute solution of an alkali sulfide or in a solution containing thiourea and sodium hydroxide.

Image Intensification and Reduction. In addition to tone changes after processing, it may also be necessary to alter the optical density or contrast of the original negative. Such alterations are rarely applied to the print. The process of increasing the image density is called intensification, whereas the process of lowering image density is called reduction. An image is intensified by adding a metal, eg, mercury, chromium, silver, or copper, to the image silver. If a given negative had sufficiently detailed information of an original scene but was underdeveloped during processing, then intensification may be useful. However, in underexposed films, the density of the original negative may not only be low but also the information may not have been recorded satisfactorily. In the latter case, intensification increases the density, but cannot provide missing details from shadow areas, ie, the toe region of the original negative (see Fig. 4). In contrast, reduction selectively removes silver from the developed image. Reduction can be used to correct overdevelopment (overexposure) or to alter the contrast of a negative. Lowering the density of a negative allows printing exposure intensities or printing exposure times to be reduced. Exposure time reduction becomes of practical importance when multiple printed copies of an original negative are required.

Intensification with mercury can be achieved by first treating the negative with a mercuric halide which oxidizes elemental silver and partly converts the silver image to a mixture of silver halide and a mercury salt. The bleached image is then darkened by a treatment with either ammonium, sodium sulfite, or a Metol-hydroquinone developer (325). During this step, both the silver and mercury ions are reduced electrochemically to their metallic forms, and thereby the image density is enhanced. An alternative single-step mercury intensification can be accomplished by using a solution containing mercuric iodide and sodium sulfite. In chromium intensification, a two-step process fundamentally equivalent to the two-step mercury process has been formulated.

Reducing treatments can be classified according to their effects on contrast. Cutting reducers reduce the density in the toe and fog regions more than in the saturation region, and therefore enhance the contrast. Reducing treatments that remove silver in proportion to the amount present at any given point in the coating are called proportional reducers. Proportional reducers lower the density of a coating with an accompanying reduction in contrast, whereas subtractive reducers decrease the density equally over the exposure scale and do not alter the contrast. In all cases, reduction involves the chemical oxidation of some of the elemental silver in the image followed by or concurrent with fixation. Ferricyanide is a commonly used oxidant, and thiosulfate can be used as the fixing agent.

Image Evaluation. The subjective quality of a developed silver image depends on the color tone of the developed silver, brightness reproduction of the original scene (35), and perceived graininess and sharpness. Certain objective measurements and analyses correlate with these subjective qualities of a developed image. For example, quantitative optical density measurements and the corresponding D-log *H* curve analysis are used to monitor tone and brightness reproduction. Quantitative and objective analysis techniques have also been developed for graininess and sharpness evaluation. The silver image is granular because it is composed of a random distribution of discrete specks and filamentary clumps of metallic silver. Consequently the photographic image is inherently inhomogeneous and nonuniform, which becomes increasingly apparent under increasing magnification. Such nonuniformity is referred to as graininess. If a given field of the silver image is optically scanned with a density measuring device, a microdensitometer,

having a sufficiently small detection aperture, spatial variation in the light-transmitting or light-reflecting properties of the image can be recorded and statistically analyzed. Such analyses produce an objective measure of graininess called granularity. If the density of a given measurement is denoted by D_i and if there are a total of N measurements in some specified uniform exposure area of the image, then the root mean square (rms) granularity, ς_D , is given by a statistical expression for the deviation in density:

$$\sigma_D = \left\{ \sum_{i=1}^N (D_i - \langle D \rangle)^2 / N \right\}^{1/2} = \{ \langle D^2 \rangle - \langle D \rangle^2 \}^{1/2}$$

where $\langle D \rangle$, the average density, is given by

$$\langle D \rangle = \sum_{i=1}^N (D_i / N)$$

The rms granularity, σ_D , is dependent on the aperture size and decreases with increasing aperture area. However, the product $A^{1/2} \sigma_D$, where A is the area of the scanning aperture, is nearly independent of aperture size in many cases (326). Accordingly, the granularity, $G = (2A)^{1/2} \varsigma_D$, called the Selwyn granularity, is often taken as a measure of the granularity of a film. Photographic films that either require high photographic speed or are ultimately to be viewed under high magnification are usually the most sensitive to the image-degrading effects of granularity. A trade-off between photographic speed and low image granularity is intrinsic to photographic films. One way to enhance the speed of a photographic film is to increase the size of the microcrystalline grains. However, grain size increases usually enhance the granular appearance of developed images.

Throughout the history of silver halide-based photography, much research has been directed toward improving the efficiency and therefore the speed without increasing the size of the silver halide grains. These research efforts have produced several improvements. Through advances in emulsion-making and sensitizing techniques, emulsion quantum efficiency has been increased. Because these advances have rendered grains of a given size more sensitive to light, the speed–grain relationship has been improved correspondingly. Furthermore, alterations in coating formats and modifications of development chemistry also have contributed significantly to improved speed–grain positions.

The visual sharpness of a recorded image is yet another subjective measure of image quality. The impression of sharpness or crispness is achieved when the boundaries and edges of the objects composing the image are clear and well defined. When high resolution of fine detail is required in the image, then, in addition to contrast and granularity, sharpness becomes a particularly important image quality parameter. The presence of a light-scattering grain population, resulting from a given set of emulsion-making conditions, can have a pronounced effect on the photographic response and the quality of the photographic image. The light-scattering and light-absorption properties of the silver halide grains within a coated layer establish the vertical and lateral distribution of light within the layer (264,327–330), which significantly affects the sharpness of the photographic image. In addition to the effects of the lateral scattering of light by emulsion grains, the optics of camera lens systems and the optics of enlargers also influence the path of photons and the apparent sharpness of the final photographic image. The ability of a photographic material to record fine detail is a function of development effects as well as of optical effects. The lateral diffusion of development by-products can produce adjacency effects that enhance the apparent sharpness of a recorded image.

Sharpness is evaluated by a number of methods. It is often measured as the ability of a recorder to produce an image of very narrow and closely spaced lines. In such an analysis, the resolving power of a recording film is determined by photographing a test object composed of a series of alternating black-and-white lines of increasing narrowness set in geometric patterns. The last visually distinguishable set of lines is recorded, in lines per millimeter, as the resolving power of the recorder under the particular test conditions. The resolving power of a photographic material is determined by granularity and contrast as well as by effects of image spread. The modulation transfer function (MTF) is a more objective and quantitatively interpretable measure of the quality of sharpness. For MTF analyses, the recorder is exposed with a band of light that varies sinusoidally in intensity. The frequency of the sine wave continuously increases along the length of the exposure band. This spatial frequency is 2–500 cycles/mm. The density modulation of a microdensitometric trace along the resultant image can be expressed as the percentage of input modulation at each frequency (34). The MTF is independent of the processing conditions for a linear recorder and can be translated mathematically through Fourier analysis to the spread function. The spread function represents the spatial distribution of light within a coating that results from the isotopic scattering of an infinitesimal incident point of light. Analysis of MTF data can also be used to deduce the image density distribution produced by knife-edge exposures; it is in general more reliable than direct analyses of such edge exposures.

BIBLIOGRAPHY

"Photography" in *ECT* 1st ed., Vol. 10, pp. 542–584, by T. H. James, Eastman Kodak Co.; in *ECT* 2nd ed., Vol. 15, pp. 355–396, by E. A. McWilliams, Eastman Kodak Co.; in *ECT* 3rd ed., Vol. 17, pp. 611–656, by D. J. Locker, Kodak Research Laboratories.

1. T. E. Whiteley, R. A. Landholm, and A. M. Kelly, *J. Soc. Sci. Photogr. Technol. Japan* **53**, 95 (1990).
2. S. Ikenoue and M. Tabei, *J. Imag. Sci.* **34**, 187 (1990).
3. M. R. Pointer and R. A. Jeffreys, *J. Photogr. Sci.* **39**, 100 (1991).
4. P. P. Koch and H. Vogler, *Ann. Phys. (Leipzig)* **77**, 495 (1925).
5. C. R. Berry, *Acta Crystallogr.* **2**, 393 (1949).
6. R. Hilsch and R. W. Pohl, *Z. Phys.* **64**, 606 (1930).
7. J. F. Hamilton, F. A. Hamm, and L. E. Brady, *J. Appl. Phys.* **35**, 414 (1964); **27**, 874 (1956).
8. E. Stenger, *The History of Photography*, Mack Printing Co., Easton, Pa., 1939, p. 1.
9. B. Newhall, *Latent Image*, Doubleday and Co., Inc., New York, 1967, p. 5.
10. *Ibid.*, p. 47.
11. *Ibid.*, p. 115.
12. *Ibid.*, p. 72.
13. S. E. Sheppard, F. H. Elliot, and S. S. Sweet, *Trans. Faraday Soc.* **19**, 261 (1923).
14. S. E. Sheppard, *Photogr. J.* **65**(N.S.49), 380 (1925).
15. Ref. 8, p. 28.
16. F. W. H. Mueller, *J. Opt. Soc. Am.* **39**, 494 (1949).
17. J. M. Harbison and J. F. Hamilton, *Photogr. Sci. Eng.* **19**, 322 (1975).
18. P. Faelens and H. Borginon, *J. Photogr. Sci.* **24**, 148 (1976).
19. J. W. Mitchell, *Photogr. Sci. Eng.* **22**, 249 (1978).
20. H. E. Spencer and A. Bunyaviroch, *J. Photogr. Sci.* **37**, 202 (1989).
21. M. K. Van Doorselaer, *J. Photogr. Sci.* **35**, 42 (1987).
22. H. W. Vogel, *The Philadelphia Photographer* **11**, 25 (1874).
23. C. B. Neblette, in J. M. Sturge, ed., *Neblette's Handbook of Photography and Reprography*, 7th ed., Van Nostrand Reinhold Co., New York, 1977, p. 8.
24. G. Binning, C. F. Quate, and C. Gerber, *Phys. Rev. Lett.* **56**, 930 (1986).
25. G. Haugstad, W. L. Gladfelter, and E. B. Weberg, *Langmuir* **9**, 3717 (1993).
26. E. H. Land, *J. Opt. Soc. Am.* **37**, 61 (1947).

27. E. H. Land, *Photogr. J.* **114**(7), 338 (1974).

28. E. H. Land, H. G. Rogers, and V. K. Walworth, in Ref. 23, p. 259; V. K. Walworth and S. H. Mervis, in J. Sturge, V. Walworth, and A. Shepp, eds., *Imaging Processes and Materials, Neblette's Eighth Edition*, Van Nostrand Reinhold, New York, 1989, p. 181.

29. W. T. Hanson, Jr., *Photogr. Sci. Eng.* **20**, 155 (1976).

30. E. H. Land, *Photogr. Sci. Eng.* **16**, 247 (1972); **21**, 225 (1977).

31. R. M. Callender, *J. Photogr. Sci.* **39**, 46 (1991).

32. C. J. V. Roberts, *J. Photogr. Sci.* **39**, 42 (1991).

33. G. C. Attridge, *J. Photogr. Sci.* **39**, 55 (1991).

34. M. A. Kriss, in T. H. James, ed., *The Theory of the Photographic Process*, 4th ed., Macmillan Publishing Co., Inc., New York, 1977, pp. 592–635.

35. C. N. Nelson, in Ref. 34, pp. 536–577.

36. E. J. Perry, *J. Colloid. Sci.* **14**, 27 (1959); **U.S. Pat. 3637391** (Jan. 25, 1972), W. Himmelmann, H. Huckstadt, R. Meyer, F. Moll, and W. Saleck (to Agfa-Gevaert); **Eur. Pat. 517961** (Dec. 16, 1992), H. Vandenabeele (to Agfa-Gevaert); **Eur. Pat. 528476** (Feb. 24, 1993), H. Vandenabeele (to Agfa-Gevaert).

37. A. W. Adamson, *Physical Chemistry of Surfaces*, John Wiley & Sons, Inc., New York, 1967, p. 337.

38. C. R. Berry, in Ref. 34, p. 88.

39. J. S. Wey, in W. R. Wilcox, ed., *Preparation and Properties of Solid State Materials*, Vol. 5, Marcel Dekker, Inc., New York, 1980.

40. G. F. Duffin, *Photographic Emulsion Chemistry*, Focal Press, London, 1966.

41. **Brit. Pat. 417,454** (Sept. 28, 1933), B. Claus.

42. **Fr. Pat. 1,173,517** (Feb. 26, 1956), C. Weaver (to E. I. du Pont de Nemours & Co., Inc.).

43. J. S. Wey, *J. Imag. Sci.* **34**, 202 (1990).

44. K. Endo and M. Okaji, *J. Photogr. Sci.* **36**, 183 (1988).

45. E. Moisar and E. Klein, *Ber. Bunsenges. Phys. Chem.* **67**, 193 (1963).

46. F. H. Claes, J. Libeer, and W. Vanassche, *J. Photogr. Sci.* **21**, 39 (1973).

47. V. K. Walworth, A. E. Rosenoff, and G. R. Bird, *Photogr. Sci. Eng.* **14**, 321 (1970).

48. W. Vanassche, F. H. Claes, H. Borginon, and J. Libeer, *J. Photogr. Sci.* **21**, 85 (1973).

49. C. R. Berry, *Phys. Rev.* **97**, 676 (1955).

50. Y. T. Tan, K. J. Lushington, P. Tangyunyong, and T. N. Rhodin, *J. Imag. Sci. Tech.* **36**, 118 (1992).

51. U. D. Schwarz, H. Haefke, H.-J. G ntherthodt, J. Bohonek, and R. Steiger, *J. Imag. Sci. Tech.* **37**, 344 (1993).

52. H. Haefke and co-workers, *J. Imag. Sci. Tech.* **37**, 545 (1993).

53. W. Markocki and A. Zaleski, *Photogr. Sci. Eng.* **17**, 289 (1973).

54. R. J. Breuhlman and F. Verhoek, *J. Am. Chem. Soc.* **70**, 1401 (1948).

55. H. Malinowski, *Z. Wiss. Photogr. Photophys. Photochem.* **51**, 186, 200 (1956).

56. H. Ch teau and J. Pouradier, *Sci. Ind. Photogr.* **24**(2), 129 (1953).

57. J. Kratochv l, B. Tezak, and V. B. Vouk, *Arh. Kem.* **26**, 191 (1954).

58. T. Tani, *J. Imag. Sci.* **32**, 227 (1988).

59. C. R. Berry, *Photogr. Sci. Eng.* **18**, 4 (1974); **20**, 1 (1976).

60. J. S. Wey and R. W. Strong, *Photogr. Sci. Eng.* **21**, 14, 248 (1977).

61. E. J. Meehan and J. K. Miller, *J. Phys. Chem.* **72**, 2168 (1968).

62. A. Hirata and S. Hohnishi, *Bull. Soc. Sci. Photogr. Jpn.* **16**, 1 (1966).

63. T. Tanaka, T. Matsubara, N. Saeki, and H. Hada, *Photogr. Sci. Eng.* **20**, 213 (1976).

64. **U.S. Pat. 2,592,250** (Apr. 8, 1952), E. P. Davey and E. B. Knott (to Eastman Kodak Co.).

65. **U.S. Pat. 3,622,318** (Nov. 23, 1971), F. J. Evans (to Eastman Kodak Co.).

66. J. Eggert and H. Fischer, *Z. Elektrochem.* **62**, 230, 393 (1958).

67. T. Shiozawa, *Bull. Soc. Photogr. Sci. Technol. Jpn.* **22**, 6 (1972).

68. C. R. Berry and S. J. Marino, *Photogr. Sci. Tech. Ser. 11*, **2**, 149 (1955).

69. B. Zuckerman, *Photogr. Sci. Eng.* **20**, 3 (1976).

70. L. J. Leyshon and P. J. Hillson, *J. Photogr. Sci.* **24**, 142 (1976).

71. C. R. Berry and D. C. Skillman, *J. Appl. Phys.* **35**, 2165 (1964).

72. J. E. Maskasky, *Photogr. Sci. Eng.* **25**, 96 (1981).

73. M. G. Antoniadis and J. S. Wey, *J. Imag. Sci. Tech.* **37**, 272 (1993).

74. Ref. 37, p. 58.

75. L. Kampmann and M. Kahlweit, *Ber. Bunsenges. Phys. Chem.* **71**, 78 (1967).

76. R. Jagannathan and J. S. Wey, *Photogr. Sci. Eng.* **26**, 61 (1982).

77. R. Jagannathan, *J. Imag. Sci.* **35**, 104 (1991).

78. I. H. Leubner, R. Jagannathan, and J. S. Wey, *Photogr. Sci. Eng.* **24**, 268 (1980).

79. I. H. Leubner, *J. Imag. Sci. Tech.* **37**, 268 (1993).

80. E. B. Gutoff, F. R. Cotrell, G. Margolis, E. G. Denk, and C. H. Wallace, *Photogr. Sci. Eng.* **18**, 8 (1974).

81. W. Ostwald, *Z. Phys. Chem. Leipzig* **34**, 495 (1900).

82. T. Sugimoto and G. Yamaguchi, *J. Colloid Interface Sci.* **57**, 400 (1976).

83. R. W. Berriman and R. H. Herz, *Nature* **180**, 293 (1957).

84. C. R. Berry and D. C. Skillman, *Photogr. Sci. Eng.* **6**, 159 (1962).

85. C. R. Berry and D. C. Skillman, *J. Photogr. Sci.* **16**, 137 (1968).

86. J. E. Maskasky, *J. Imag. Sci.* **31**, 15 (1987).

87. R. V. Mehta, R. Jagannathan, and J. A. Timmons, *J. Imag. Sci. Tech.* **37**, 107 (1993).

88. T. Sugimoto, *Photogr. Sci. Eng.* **28**, 137 (1984).

89. P. H. Karpinski and J. S. Wey, *J. Imag. Sci.* **32**, 34 (1988).

90. T. Sugimoto, *J. Imag. Sci.* **33**, 203 (1989).

91. T. Tani and M. Saito, *Photogr. Sci. Eng.* **17**, 434 (1973).

92. J. Riesenfeld, *Photogr. Sci. Eng.* **17**, 213 (1973).

93. H. Hediger and P. Junod, *Photogr. Sci. Eng.* **20**, 50 (1976).

94. J. VanBiesen, *J. Appl. Phys.* **41**, 1910 (1970).

95. H. A. Hoyen, Jr., *J. Appl. Phys.* **47**, 3784 (1976).

96. S. Takada, *Photogr. Sci. Eng.* **18**, 500 (1974).

97. P. Junod, E. Schumacker, R. Steiger, and J. Wullschleger, *Photogr. Sci. Eng.* **17**, 205 (1973).

98. J. F. Hamilton and L. E. Brady, *J. Appl. Phys.* **30**, 1893 (1959).

99. A. P. Batra, A. L. Laskar, and L. Slifkin, *J. Phys. Chem. Solids* **30**, 2053 (1969).

100. P. S ptitz and R. Weidman, *Phys. Status Solidi* **27**, 631 (1968).

101. C. Bucci and co-workers, *J. Phys. Chem. Solids* **23**, 363 (1965).

102. C. Bucci, *Phys. Rev.* **164**, 1200 (1967).

103. A. P. Batra and L. M. Slifkin, *Photogr. Sci. Eng.* **17**, 64 (1973).

104. M. Hühne and M. Stasiw, *Phys. Status Solidi* **20**, 657 (1967).

105. R. Eachus and R. E. Graves, *J. Chem. Phys.* **63**, 83 (1975); **65**, 5445 (1977).

106. Y. T. Tu and J. M. Blakely, *Photogr. Sci. Eng.* **20**, 59 (1976).

107. S. H. Ehdich and I. H. Leubner, *J. Imag. Sci. Tech.* **36**, 105 (1992).

108. B. H. Carroll, *Photogr. Sci. Eng.* **24**, 265 (1980).

109. R. S. Eachus and M. T. Olm, *Cryst. Latt. Def. Amorph. Mater.* **8**, 297 (1989).

110. M. Todorova, A. Stankova, I. Markov, and M. Pacheva, *J. Photogr. Sci.* **35**, 196 (1987).

111. R. S. Eachus and J. P. Spoonhower, *Photogr. Sci. Eng.* **26**, 20 (1982).

112. M. Todorova, I. Markov, A. Stankova, and M. Karabelova, *J. Photogr. Sci.* **34**, 91 (1986).

113. D. Yin, J. Yue, T.-t. Yan, and B.-x. Peng, *J. Photogr. Sci.* **39**, 11 (1991).

114. M. R. V. Sahyun, *Photogr. Sci. Eng.* **27**, 171 (1983).

115. V. Weiss, D. Wyrsh, and M. Zobrist, *Photogr. Sci. Eng.* **27**, 40 (1983).

116. A. Hoffman, *Photogr. Sci. Eng.* **27**, 261 (1983).

117. K. Endo and M. Saikawa, *J. Photogr. Sci.* **38**, 210 (1990).

118. M. T. Olm, J. R. Niklas, J. M. Spaeth, and M. C. R. Symons, *Phys. Rev. B.* **38**, 4343 (1988).

119. J. W. Mitchell, *J. Imag. Sci.* **31**, 239 (1987).

120. Y. B. Peng, G. Vekeman, W. Maenhout-Van der Vorst, and F. Cardon, *J. Imag. Sci.* **31**, 55 (1987).

121. R. S. Eachus and M. T. Olm, *Radiation Effects* **73**, 69 (1983).

122. L. de Brabandere and P. Faelens, *Photogr. Sci. Eng.* **25**, 63 (1981).

123. **U.S. Pat. 3,118,766** (Jan. 21, 1964), P. H. Roth (to Polaroid Corp.).

124. **U.S. Pat. 2,614,928** (Oct. 21, 1952), H. C. Yutzy and G. F. Frame (to Eastman Kodak Co.).

125. **U.S. Pat. 2,614,929** (Oct. 21, 1952), H. C. Yutzy and F. J. Russell (to Eastman Kodak Co.).

126. **U.S. Pat. 3,218,169** (Nov. 16, 1975), W. D. Kelly, P. H. Roth, and L. D. Taylor (to Polaroid Corp.).

127. D. Denk, *Photogr. Sci. Eng.* **20**, 35 (1976).

128. **U.S. Pat. 3,884,701** (May 20, 1975), A. J. Depauw and J. A. Carpentier (to Agfa-Gevaert, Inc.).

129. **U.S. Pat. 5,164,092** (Nov. 17, 1992); **U.S. Pat. 5,248,418** (Sept. 28, 1993), W. D. Munch (to Eastman Kodak Co.); **U.S. Pat. 4,334,012** (June 8, 1982), A. G. E. Mignot (to Eastman Kodak Co.); **U. S. Pat. 4,336,328** (June 22, 1982), B. M. Brown, M. L. Judd, and R. N. Schindler (to Eastman Kodak Co.); **U.S. Pat. 5,242,597** (Sept. 7, 1993), J. M. McArdle (to Eastman Kodak Co.); **U.S. Pat. 4,758,505** (July 19, 1988), K. Hoffmann (to Agfa Gevaert Aktiengesellschaft).

130. J. W. Mitchell, *Photogr. Sci. Eng.* **26**, 270 (1982); *J. Imag. Sci. Tech.* **37**, 331 (1993).

131. M. Bucher, *J. Imag. Sci.* **34**, 89 (1990).

132. C. Kittel, *Introduction to Solid State Physics*, John Wiley & Sons, Inc., New York, 1971, Chapt. 9.

133. C. R. Berry, *Photogr. Sci. Eng.* **19**, 93 (1975).

134. F. Granzer, *J. Imag. Sci.* **33**, 207 (1989).

135. C. R. Berry, *J. Photogr. Sci.* **18**, 169 (1970).

136. H. Kanzaki, *Photogr. Sci. Eng.* **24**, 219 (1980).

137. J. W. Mitchell, *Sci. Ind. Photogr.* **28**, 493 (1957).

138. J. W. Mitchell, *J. Phys. Chem.* **66**, 2359 (1962).

139. H. E. Spencer, L. E. Brady, and J. F. Hamilton, *J. Opt. Soc. Am.* **54**, 492 (1964).

140. J. F. Hamilton, in J. O. McCaldin, ed., *Progress in Solid State Chemistry*, Vol. 8, Pergamon Press, Inc., New York, 1973, p. 167.

141. K. Ohzeki, S. Urabe, and T. Tani, *J. Imag. Sci.* **34**, 136 (1990).

142. J. F. Hamilton, *Photogr. Sci. Eng.* **18**, 493 (1974).

143. **U.S. Pat. 4,665,612** (1988), T. Sugimoto.

144. **U.S. Pat. 4,668,614** (1988), S. Takada.

145. **U.S. Pat. 4,433,048** (1984), J. C. Solberg, R. H. Pigginn, and H. S. Wilgus (to Eastman Kodak Co.).

146. **U.S. Pat. 4,400,463** (1983) and **U.S. Pat. 4,435,501** (1984), J. E. Maskasky (to Eastman Kodak Co.).

147. S. Bando, Y. Shibahara, and S. Ishimaru, *J. Imag. Sci.* **29**, 193 (1985).

148. Y. A. Beslav and L. V. Kolesnikov, *J. Photogr. Sci.* **39**, 2 (1991).

149. P. Bi-xian, W. Rony-qin, L. Zhen-xing, Z. Hong, and Z. Xiao-hong, *J. Photogr. Sci.* **41**, 2 (1993).

150. N. B. Hannay, *Solid-State Chemistry*, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1967, pp. 80–98.

151. R. B. Poepfel and J. M. Blakely, *Surf. Sci.* **15**, 507 (1969).

152. Y. T. Tan and H. A. Hoyen, Jr., *Surf. Sci.* **36**, 242 (1973).

153. H. A. Hoyen, *Photogr. Sci. Eng.* **17**, 188 (1973).

154. S. Takada, *Photogr. Sci. Eng.* **19**, 214 (1975).

155. R. A. Hudson, S. K. Wonnell, G. C. Farlow, and L. M. Slifkin, *J. Imag. Sci.* **34**, 101 (1990).

156. J. W. Mitchell, *Photogr. Sci. Eng.* **23**, 1 (1979); **22**, 1 (1978); **25**, 170 (1981); *J. Photogr. Sci.* **31**, 222 (1983); *J. Imag. Sci.* **34**, 113 (1990); **34**, 217 (1990); **35**, 341 (1991).

157. N. F. Mott, *Photogr. J.* **81**, 62 (1941).

158. R. W. Gurney and N. F. Mott, *Proc. R. Soc. London* **164**, 151 (1938).

159. J. F. Hamilton, *Photogr. Sci. Eng.* **14**, 102 (1970); **12**, 143 (1968).

160. M. R. V. Sahyun, *J. Imag. Sci.* **33**, 94 (1989).

161. R. K. Hailstone, N. B. Liebert, M. Levy, and J. F. Hamilton, *J. Photogr. Sci.* **36**, 2 (1988); *J. Imag. Sci.* **31**, 185 (1987).

162. R. K. Hailstone and J. F. Hamilton, *J. Imag. Sci.* **29**, 125 (1985); *J. Photogr. Sci.* **34**, 2 (1986).

163. R. C. Baetzold, *J. Photogr. Sci.* **28**, 15 (1980).

164. E. Moisar, *Photogr. Sci. Eng.* **26**, 124 (1982).

165. E. Moisar, F. Granzer, D. Dautrich, and E. Palm, *J. Photogr. Sci.* **28**, 71 (1980).

166. A. Panov and J. Malinowski, *J. Photogr. Sci.* **33**, 75 (1985).

167. N. Starbova and J. Malinowski, *J. Photogr. Sci.* **32**, 245 (1985).

168. A. Kempf, *Z. Wiss. Photogr.* **36**, 235 (1937).

169. D. J. Locker, *Photogr. Sci. Eng.* **18**, 242 (1974); **19**, 160 (1975).

170. R. L. Jenkins and G. C. Farnell, *J. Photogr. Sci.* **20**, 167 (1972).

171. S. F. Sheppard, A. P. H. Trivelli, and R. P. Loveland, *J. Franklin Inst.* **200**, 51 (1925).

172. S. E. Sheppard, *Photogr. J.* **68**, 397 (1928).

173. R. A. Bassett and H. O. Dickinson, *J. Photogr. Sci.* **11**, 239 (1963).

174. R. Venet, H. Chsteau, and J. Pouradier, *Sci. Ind. Photogr.* **29**, 220 (1948).

175. H. Chsteau and J. Pouradier, *Sci. Ind. Photogr.* **31**, 262 (1960).

176. D. M. Spracklen, *J. Photogr. Sci.* **14**, 220 (1966); **15**, 249 (1967).
177. D. J. Cash, *J. Photogr. Sci.* **16**, 102 (1968).
178. M. Ridgway and P. J. Hillson, *J. Photogr. Sci.* **23**, 153 (1975).
179. R. Koslowsky and H. Mueller, *Z. Wiss. Photogr. Photophys. Photochem.* **46**, 65 (1951).
180. G. C. Farnell and L. R. Solman, *J. Photogr. Sci.* **24**, 203 (1976).
181. T. A. Babcock, P. M. Ferguson, W. C. Lewis, and T. H. James, *Photogr. Sci. Eng.* **19**, 49, 211 (1975).
182. D. J. Locker, *Photogr. Sci. Eng.* **17**, 442 (1973).
183. G. A. Janusonis, *Photogr. Sci. Eng.* **22**, 297 (1978).
184. L. F. Costa, G. A. Janusonis, and J. A. Merrigan, *Photogr. Sci. Eng.* **22**, 301 (1978).
185. D. J. Cash, *J. Photogr. Sci.* **29**, 140 (1981).
186. E. Moisar, *J. Photogr. Sci.* **14**, 220 (1966); **15**, 249 (1967).
187. D. J. Cash, *J. Photogr. Sci.* **20**, 107 (1972).
188. K. V. Chibisov, A. A. Titov, and A. A. Mikhailova, *Zh. Fiz. Khim.* **27**, 866 (1953).
189. D. Corbin, A. Gingello, G. MacIntyre, and B. H. Carroll, *Photogr. Sci. Eng.* **24**, 45 (1980).
190. J. E. Keevert and V. V. Gokhale, *J. Imag. Sci.* **31**, 243 (1987).
191. L. M. Kellogg, N. B. Liebert, and T. H. James, *Photogr. Sci. Eng.* **16**, 115 (1970).
192. D. Zhang and R. K. Hailstone, *J. Imag. Sci. Tech.* **37**, 61 (1993).
193. E. Fatuzzo and S. Coppo, *J. Appl. Phys.* **42**, 4752 (1971).
194. W. West and V. I. Saunders, *Photogr. Sci. Eng.* **3**, 258 (1959).
195. E. Moisar, *Photogr. Sci. Eng.* **25**, 45 (1981).
196. E. Moisar and F. Granzer, *Photogr. Sci. Eng.* **26**, 1 (1982).
197. S. Sonoike and T. Nagashima, *J. Imag. Sci.* **33**, 227 (1989).
198. J. Dikova and K. Starbova, *J. Imag. Sci.* **33**, 91 (1989).
199. S. H. Ehrlich, *J. Imag. Sci. Tech.* **37**, 73 (1993).
200. F. W. H. Mueller, *Photogr. Sci. Eng.* **10**, 338 (1966).
201. H. E. Spencer and R. E. Atwell, *J. Opt. Soc. Am.* **54**, 498 (1964); **56**, 1095 (1966).
202. D. A. Pitt, M. L. Rachu, and M. R. V. Sahyun, *Photogr. Sci. Eng.* **25**, 57 (1981).
203. A. E. Steigmann, *Sci. Ind. Photogr.* **22**, 441 (1951).
204. H. W. Wood, *J. Photogr. Sci.* **6**, 33 (1958); **1**, 163 (1953).
205. A. Hautot and H. Sauvenier, *Sci. Ind. Photogr.* **24**, 257 (1953).
206. Yu. Sh. Mushkovskii, *Zh. Nauchn. Prikl. Fotogr. Kinematogr.* **3**, 51 (1958).
207. R. K. Hailstone, N. B. Liebert, M. Levy, and J. F. Hamilton, *J. Imag. Sci.* **35**, 219 (1991).
208. S. S. Collier, *Photogr. Sci. Eng.* **23**, 113 (1979).
209. T. Tani and S. Takada, *Photogr. Sci. Eng.* **26**, 111 (1982).
210. J. F. Hamilton and R. C. Baetzold, *Photogr. Sci. Eng.* **25**, 189 (1981).
211. **U.S. Pat. 4,433,048** (1989), J. C. Solberg, R. H. Piggan, and H. S. Wilgus (to Eastman Kodak Co.).
212. A. H. Herz, in Ref. 34, pp. 235–250.
213. T. Tani, *J. Imag. Sci.* **29**, 165 (1985).
214. D. D. F. Shiao, *J. Imag. Sci.* **30**, 22 (1986).
215. C. Reich, W. D. Pandolfe, and G. R. Bird, *Photogr. Sci. Eng.* **17**, 334 (1973).
216. S. Dahne, *J. Imag. Sci. Tech.* **38**, 101 (1994).
217. D. Mastropaolo, J. Potenza, and G. R. Bird, *Photogr. Sci. Eng.* **18**, 450 (1974).
218. W. West and P. B. Gilman, in Ref. 34, pp. 251–290.
219. D. M. Sturmer, W. S. Gaugh, and B. J. Bruschi, *Photogr. Sci. Eng.* **18**, 49, 56 (1974); D. M. Sturmer and A. P. Marchetti, in Ref. 28.
220. J. Lenhard and R. L. Parton, *J. Am. Chem. Soc.* **109**, 5808 (1987).
221. T. H. James, *Photogr. Sci. Eng.* **16**, 120 (1972).
222. J. W. Mitchell, *J. Photogr. Sci.* **40**, 94 (1992); **31**, 185 (1983).
223. P. B. Gilman, *Photogr. Sci. Eng.* **18**, 418 (1974); *J. Photogr. Sci.* **31**, 185 (1983).
224. J. Siegel, D. Fassler, M. Friedrich, J. V. Grossmann, U. Kempka, and H. Pietsch, *J. Photogr. Sci.* **35**, 73 (1987).
225. T. Tani, *Photogr. Sci. Eng.* **26**, 213 (1982); *J. Imag. Sci.* **33**, 17 (1989); **34**, 143 (1990); *J. Appl. Phys.* **62**, 2456 (1987); T. Tani, T. Suzumoto, K. Kemnitz, and K. Yoshihara, *J. Phys. Chem.* **96**, 2778 (1992).
226. I. H. Leubner, *Photogr. Sci. Eng.* **18**, 175 (1974); **20**, 61 (1976); **24**, 138 (1980).
227. R. K. Hailstone, *J. Photogr. Sci.* **32**, 25 (1984).
228. D. M. Sturmer and D. W. Heseltine, in Ref. 34, pp. 194–234.
229. L. G. S. Brooker, R. H. Sprague, C. P. Smyth, and G. L. Lewis, *J. Am. Chem. Soc.* **62**, 1116 (1940).
230. J. H. Noggle, *Physical Chemistry*, Scott, Foresman and Co., Boston, Mass., 1989, pp. 633–639.
231. **U.S. Pat. 2,756,227** (July 24, 1956); **2,734,900** (Feb. 14, 1956), D. W. Heseltine (to Eastman Kodak Co.).
232. J. R. Lenhard, B. R. Hein, and A. A. Muentner, *J. Phys. Chem.* **97**, 8269 (1993).
233. W. West and A. L. Geddes, *J. Chem. Phys.* **68**, 837 (1964).
234. J. Eggers, E. G  nther, and E. Moisar, *Photogr. Korresp.* **102**, 144 (1966).
235. J. R. Lenhard and A. A. Muentner, in D. S. Ginley, ed., *Photoelectrochemistry and Electrosynthesis on Semiconducting Materials* Proceedings, Vol. 88-14, The Electrochemical Society, Pennington, N.J., 1988, pp. 97–104.
236. **U.S. Pat. 2,131,038** (1938), L. G. S. Brooker and C. J. Staud (to Eastman Kodak Co.).
237. **U.S. Pat. 2,694,716** (1955), C. F. H. Allen and V. Wilson.
238. E. J. Birr, *Z. Wiss. Photogr.* **50**, 107 (1955).
239. E. J. Birr, *Stabilization of Photographic Silver Halide Emulsions*, Focal Press, Inc., New York, 1974.
240. M. R. V. Sahyun, *Photogr. Sci. Eng.* **18**, 383 (1974).
241. T. Tani, *Photogr. Sci. Eng.* **25**, 230 (1981).
242. Z. C. H. Tan, *Photogr. Sci. Eng.* **19**, 17 (1975).
243. H. D. Meissner, *J. Photogr. Sci.* **19**, 83 (1971).
244. T. Tani, *Photogr. Sci. Eng.* **22**, 179 (1978).
245. T. Tani, *Photogr. Sci. Eng.* **19**, 114 (1975).
246. E. J. Birr, *Z. Wiss. Photogr.* **47**, 2; **48**, 103 (1953).
247. D. D. F. Shiao, M. T. Nieh, and A. H. Herz, *J. Photogr. Sci.* **30**, 208 (1982).
248. D. L. Smith and H. R. Luss, *Photogr. Sci. Eng.* **20**, 184 (1976).
249. **Br. Pat. 636,234** (1939), J. A. Leemakers, E. C. Yakel, and E. H. H. Hewitson (to Eastman Kodak Co.).
250. D. M. Burness and J. Pouradier, in Ref. 34, pp. 77–84.
251. A. M. Venet and J. Pouradier, *Bull. Soc. Chim. Fr.* 1336 (1955).

252. S. K. Taylor, F. Davidson, and D. W. Overall, *Photogr. Sci. Eng.* **22**, 134 (1978).
253. P. Kowaliski, *Applied Photographic Theory*, John Wiley & Sons, Inc., New York, 1972, pp. 343–354.
254. T. F. Parsons, G. G. Gray, and I. H. Crawford, *J. Appl. Photogr. Eng.* **5**, 110 (1979).
255. V. L. Zelikman, *Zh. Nauchn. Prikl. Fotogr. Kinematogr.* **12**, 379 (1967).
256. V. L. Zelikman and S. M. Levi, in E. A. Sutherns, ed., *Making and Coating Photographic Emulsions*, Focal Press Inc., New York, 1964, pp. 277–306.
257. E. Cohen and E. Gutoff, *Modern Coating and Drying Technology*, VCH Publishers, Inc., New York, 1992.
258. F. W. H. Mueller, in Ref. 23, p. 28.
259. R. Audran, *Sci. Ind. Photogr.* **27**, 433 (1956).
260. Ref. 253, p. 236.
261. **Brit. Pat. 635,841** (Apr. 19, 1950), E. P. Davey and E. B. Knott (to Kodak Ltd.).
262. **U.S. Pat. 3,206,313** (Sept. 14, 1965); **3,317,322** (May 2, 1967), H. D. Porter, T. H. James, and W. G. Lowe (to Eastman Kodak Co.).
263. **U.S. Pat. 3,367,778** (Feb. 6, 1968), R. W. Berriman (to Eastman Kodak Co.).
264. J. J. DePalma and J. Gasper, *Photogr. Sci. Eng.* **16**, 181 (1972).
265. G. S. Allbright, *J. Photogr. Sci.* **39**, 95 (1991).
266. R. E. Bacon, in Ref. 34, pp. 182–193.
267. J. A. Aznórez, *Photogr. Sci. Eng.* **23**, 248 (1979).
268. S. Sakai and S. Baba, *Bull. Soc. Sci. Photogr. Jpn.* **17**, 12 (1967).
269. **U.S. Pat. 2,717,833** (Sept. 13, 1955), H. Wark (to Sperry Rand Corp.).
270. **U.S. Pat. 3,501,305** (Mar. 17, 1970), R. W. Berriman (to Eastman Kodak Co.).
271. R. C. Baetzold, *Photogr. Sci. Eng.* **17**, 78 (1973).
272. J. D. Kendall, in L. P. Clerc, ed., *IX^e Congrès International de Photographie Scientifique et Appliquée, Paris 1935*, Editions Revue d'Optique, Paris, 1936, p. 227.
273. H. W. Altland, G. E. Smith, I. A. Olivares, and S. W. Cowan, *Photogr. Sci. Eng.* **23**, 257 (1979).
274. W. E. Lee and E. R. Brown, in Ref. 34, pp. 291–334.
275. R. B. Pontius, R. G. Willis, and R. J. Newmiller, *Photogr. Sci. Eng.* **16**, 406 (1972).
276. D. D. F. Shiao and T. Bistrovich, *J. Imag. Sci.* **35**, 279 (1991).
277. J. Belloni, M. Mostafavi, J. L. Marignier, and J. Amblard, *J. Imag. Sci.* **35**, 68 (1991).
278. D. D. F. Shiao, *J. Imag. Sci.* **31**, 64, 70 (1987).
279. G. I. P. Levenson, *J. Photogr. Sci.* **39**, 63 (1991).
280. R. B. Pontius and R. G. Willis, *Photogr. Sci. Eng.* **17**, 21, 326 (1973).
281. J. O'M. Bockris, N. Bonciocat, and F. Gutmann, *An Introduction to Electrochemical Science*, Springer-Verlag New York, Inc., New York, 1974, p. 63.
282. T. H. James, in Ref. 34, pp. 373–432.
283. W. Jaenicke, *J. Photogr. Sci.* **20**, 2 (1972).
284. E. R. Brown and L. K. J. Tong, *Photogr. Sci. Eng.* **19**, 314 (1975).
285. I. Konstantinov, A. Panov, and J. Malinowski, *J. Photogr. Sci.* **21**, 250 (1973).
286. I. Konstantinov and J. Malinowski, *J. Photogr. Sci.* **23**, 1 (1975).
287. P. J. Hillson, *J. Photogr. Sci.* **22**, 31 (1974).
288. J. H. W. Cramp and P. J. Hillson, *J. Photogr. Sci.* **24**, 30 (1976).
289. W. E. Mueller, *Photogr. Sci. Eng.* **15**, 369 (1971); **17**, 94 (1973).
290. Zh. L. Broun and K. V. Chibisov, *Z. Wiss. Photogr.* **57**, 166 (1963).
291. W. Gross, *J. Photogr. Sci.* **19**, 207 (1971).
292. G. Eberhard, *Phys. Z.* **13**, 288 (1912).
293. J. A. C. Yule, *J. Franklin Inst.* **239**, 221 (1945).
294. R. E. Stauffer, W. F. Smith, and A. P. H. Trivelli, *J. Franklin Inst.* **238**, 291 (1944).
295. H. D. Hunt, *Photogr. Sci. Eng.* **5**, 104 (1961).
296. J. W. Carpenter and P. W. Lauf, *Research Disclosure*, Item 17029, June 1978, pp. 9–15.
297. M. S. Borodkina, S. A. Nedzhii, V. I. Sheberstov, and O. A. Yatskevich, *Zh. Nauchn. Prikl. Fotogr. Kinematogr.* **12**(2), 127 (1967).
298. A. G. Leiga, *J. Phys. Chem.* **70**, 3260 (1966).
299. D. F. Schultze, *J. Appl. Photogr. Eng.* **5**, 163 (1979).
300. J. Van den Houte and L. Vermeulen, *J. Appl. Photogr. Eng.* **5**, 167 (1979).
301. A. Rott and E. Weyde, *Photographic Silver Halide Diffusion Process*, Focal Press, Inc., New York, 1972.
302. G. I. P. Levenson, in Ref. 34, p. 437.
303. G. I. P. Levenson and M. G. Rumens, *J. Photogr. Sci.* **13**, 79 (1965).
304. A. Green, *Photogr. Sci. Eng.* **19**, 124 (1975).
305. Ref. 250, pp. 487–494.
306. **U.S. Pat. 2,875,048** (Feb. 24, 1959), G. M. Haist and J. R. King (to Eastman Kodak Co.).
307. H. D. Russell, E. C. Yackel, and J. S. Bruce, *PSA J. Sect. B* **16**, 59 (1950).
308. H. D. Russell, *Photogr. Eng.* **2**, 136 (1951).
309. L. Wolfman, *1991–92 Wolfman Report on the Photographic Industry in the United States*, Hachette Magazines, Inc., New York, 1992.
310. *U-1000 Kodak Literature Packet*, Department 412-L, Eastman Kodak Co., Rochester, New York.
311. D. A. Branch, *J. Imag. Tech.* **14**, 160 (1988).
312. T. W. Bober and T. J. Dagon, *Image Technol.* **14**(4), 13 (1972); **14**(5), 19 (1972).
313. A. Cooley and D. Vacco, *J. Imag. Sci. Tech.* **37**, 603 (1993).
314. R. E. Jacobson, *J. Photogr. Sci.* **39**, 70 (1991).
315. K. B. Hendriks, in J. Sturge, V. Walworth and A. Shepp, ed., *Imaging Processes and Materials, Neblette's Eighth Edition*, Van Nostrand Reinhold, New York, 1989, p. 637.
316. R. W. Henn and D. G. Wiest, *Photogr. Sci. Eng.* **7**, 253 (1963).
317. R. W. Henn, D. G. Wiest, and B. D. Mack, *Photogr. Sci. Eng.* **9**, 167 (1965).
318. W. E. Lee and C. C. Bard, *J. Photogr. Sci.* **36**, 73 (1988).
319. D. C. Skillman, *Photogr. Sci. Eng.* **19**, 28 (1975).
320. R. B. Flint and G. C. Farnell, *J. Photogr. Sci.* **21**, 175 (1973).
321. H. D. Keith and J. W. Mitchell, *Philos. Mag.* **44**, 877 (1953).
322. W. Jaenicke, *Photogr. Sci. Eng.* **6**, 185 (1962).
323. P. B. Price, D. A. Vermilyea, and M. B. Webb, *Acta Metall.* **6**, 524 (1958).
324. E. Weyde, E. Klein, and H. J. Metz, *J. Photogr. Sci.* **10**, 110 (1962).
325. J. S. Carroll, *Photographic Lab Handbook*, 4th ed., American Photographic Book Publishing Co., Inc., Garden City, N.Y., 1977, pp. 190–194.
326. E. W. H. Selwyn, *Photogr. J.* **75**, 571 (1935); **82**, 208 (1942).

327. J. Gasper and J. J. DePalma, in Ref. 34, Chapt. 20, p. 578.
328. E. Pitts, *Proc. Phys. Soc. London* **67B**, 105 (1954).
329. P. Kubelka and F. Munk, *Z. Techn. Phys.* **12**, 593 (1931).
330. P. S. Mudgett and L. W. Richards, *Appl. Opt.* **10**, 1485 (1971).

David J. Locker
Kodak Manufacturing Research and Engineering



PHOTOMULTIPLIER TUBES.

See PHOTODETECTORS.

PHOTOREACTIVE POLYMERS.

See RESIST MATERIALS.

PHOTORESISTS.

See RESIST MATERIALS.

PHOTOVOLTAIC CELLS

A photovoltaic (PV) solar power system is a complete electrical source that uses solar cells to directly convert light energy into electricity. The system can be self-contained and completely autonomous or it can work in tandem with other conventional fuel-based sources of power to offer robust power availability.

A solar cell is a semiconductor device that can convert light instantaneously into direct-current (d-c) electricity. A number of cells are typically connected together in series in a weather-resistant package such that enough voltage is generated to recharge a 12-volt lead–acid storage battery, the most common storage device used in conjunction with solar power (see BATTERIES, LEAD ACID). Such a package of cells is designated a PV module, which is often constructed of an external sheet of strengthened glass and polymeric encapsulation. The most common size module is 0.5–1 m² in area and delivers between 25 and 150 watts of power.

The advantages of photovoltaic cells as a source of electric power over alternative power sources may be characterized as follows: solar cells capture sunlight, an essentially inexhaustible and nonpolluting energy source which is freely distributed, and directly convert that light into electricity; photovoltaic generation of electricity requires no machinery with moving parts and produces no noise, waste, or polluting by-products; photovoltaic systems are modular and therefore can be adapted for a variety of applications. Solar power systems are particularly useful in areas where power lines cannot be readily or inexpensively routed.

Solar cells have been used extensively and successfully to power satellites in space since the late 1950s, where their high power-to-weight ratio and demonstrated reliability are especially desirable characteristics. On earth, where electrical systems typically provide large amounts of power at reasonable costs, three principal technical limitations have thus far impeded the widespread use of photovoltaic products: solar cells are expensive, sunlight has a relatively low power density, and commercially available solar cells convert sunlight to electricity with limited efficiency. Clearly, terrestrial solar cells must be reasonably efficient, affordable, and durable. International efforts are dedicated to obtaining such devices, and a number of these activities have been reviewed (1).

The power density of sunlight is about 1350 W m² at elevations just above the earth's atmosphere (2). Less than 1000 W m² is typically incident on earth after filtering through the atmosphere. For comparison, in every four clear daylight hours on earth, energy from the sun that falls on an area the size of a large car is equivalent to that stored in 3.8 L (~1 gal) of gasoline. That amount of gasoline will produce about 10 kWh of electrical output if consumed in a conventional power plant. The conversion efficiencies, which theoretically can be attained from the most efficient cell materials, are 25–40% under standardized test conditions, but typical cell performance falls below these limits. The conversion efficiency of a solar cell is the ratio of electric power output to solar power input. Due to the low power density of sunlight and limited conversion efficiencies, the most efficient solar modules can generate about 250 W m² in peak sunlight conditions. The maximum power output of a solar cell or module is defined in peak watts (W_{peak}), a rating based on a standard measurement method established by international consensus. A solar panel of one square meter area nominally produces one kilowatt hour of electricity per day. For most large-scale, power-producing applications, solar modules have conversion efficiencies above 10% in order to minimize the total cost of a generating system.

Chemistry

Crystalline silicon *p–n* junction solar cells are the principal commercially available type and are used here to illustrate the operation of a solar cell. When sunlight falls on a solar cell, a voltage is induced and an electric current flows in an external circuit that is connected to the cell. Each atom in the silicon crystal lattice is surrounded by and bound to four equidistant neighboring atoms. The outermost shell of electrons of each silicon atom contains four valence electrons, and each of the four valence electrons in the crystal lattice is shared in a bonding orbital with an electron from one of its four nearest neighbors. This electron pair or covalent bond firmly binds the crystal. If all the valence electrons were inexorably bound, as they would be at 0 degrees kelvin, the silicon crystal would be an insulator because no free electrons would be available, and conduction would be precluded (Fig. 1a). However, the covalent bonds can be broken, eg, by thermal excitation (Fig. 1b). The energy required to break a covalent bond is the bond energy or energy gap, *E_g*. In silicon, *E_g* is ca 1.1 eV.

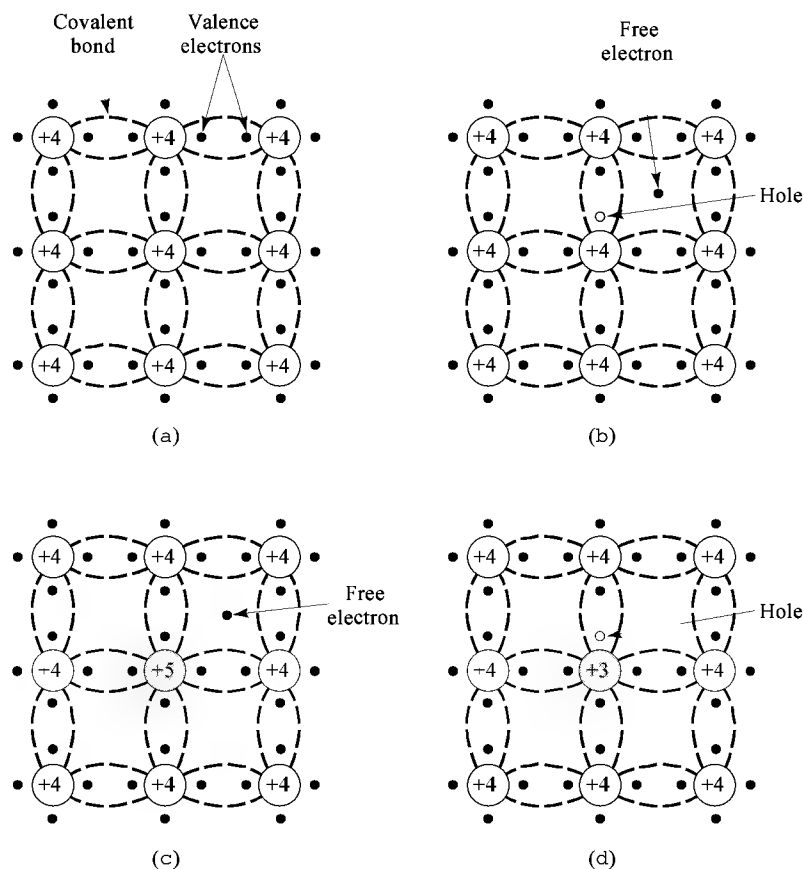


Fig. 1. (a) Silicon (valence = 4) crystal lattice shown in two dimensions with no broken bonds, $T = 0\text{ K}$; (b) silicon crystal lattice with a broken bond; (c) silicon crystal lattice with a silicon atom displaced by a donor dopant, ie, n -doped (valence = 5); and (d) silicon crystal lattice with a silicon atom displaced by an acceptor dopant, ie, p -doped (valence = 3).

The absence of an electron from a covalent bond leaves a hole and the neighboring valence electron can vacate its covalent bond to fill the hole, thereby creating a hole in a new location. The new hole can, in turn, be filled by a valence electron from another covalent bond, and so on. Hence, a mechanism is established for electrical conduction that involves the motion of valence electrons but not free electrons. Although a hole is a conceptual artifact, it can be described as a concrete physical entity to keep track of the motion of the valence electrons. Because holes and electrons move in opposite directions under the influence of an electric field, a hole has the same magnitude of charge as an electron but is opposite in sign.

The energy in light also can break the bonds of silicon valence electrons. Each photon has energy equal to the product of Planck's constant and the frequency of the light, ie, $E = h\nu$, where E is photon energy, h is Planck's constant, and ν is the frequency of light. Solar photons range in energy from 0.5 eV for infrared to 4 eV for ultraviolet.

When a photon having energy equal to or greater than E_g is absorbed by the silicon crystal, the photon breaks a covalent bond, thereby freeing an electron and forming a hole. An electron is excited by a photon from a valence-energy band in a covalent bond into a conduction-energy band. The electron, which is transformed into a mobile, negatively charged carrier, leaves behind a mobile hole and consequently the photon has formed a free electron-hole pair.

If the hole and electron are not kept apart, they recombine to produce a small amount of thermal energy within the crystal and no net current flow. When the holes and electrons are kept apart, collected, and made to flow in a circuit outside the crystal, they produce electric current in that circuit. Solar cells are equipped with a barrier or a junction which provides an internal electric field that segregates photogenerated electrons and holes. Thus, although unmodified silicon has an equal number of holes and electrons, a p - n junction silicon solar cell consists of two charge-dissimilar regions which are separated by a junction: one region is rich in holes (positive), ie, p -type silicon, and the other is rich in electrons (negative), ie, n -type silicon. Such regions do not occur naturally; they are fabricated by doping, ie, replacing some silicon atoms in the lattice with atoms having a valence other than four. Replacement of a few silicon atoms, ie, ca one in several million, causes large increases in the electrical conductivity of the resultant doped crystal.

Atoms of elements that are characterized by a valence greater than four, eg, phosphorus or arsenic (valence = 5), are one type of dopant. These high valence dopants contribute free electrons to the crystal and are called donor dopants. If one donor atom is incorporated in the lattice, four of the five valence electrons of donor dopants are covalently bonded, but the fifth electron is very weakly bound and can be detached by only ca 0.03 eV of energy. Once it is detached, it is available as a free electron, ie, a carrier of electric current. A silicon crystal with added donor dopants has excess electron carriers and is called n -type (negative) silicon (Fig. 1c).

When a silicon crystal is doped with atoms of elements having a valence of less than four, eg, boron or gallium (valence = 3), only three of the four covalent bonds of the adjacent silicon atoms are occupied. The vacancy at an unoccupied covalent bond constitutes a hole. Dopants that contribute holes, which in turn act like positive charge carriers, are acceptor dopants and the resulting crystal is p -type (positive) silicon (Fig. 1d).

Conductivity in doped silicon crystals is determined by the properties of the added charge carriers or majority carriers. In n -type silicon, electrons are majority carriers and holes are minority carriers. There are fewer holes in n -type silicon than in undoped silicon because the large number of electrons causes some recombination with preexisting holes. In p -type silicon, holes are the majority carriers and electrons are the minority carriers. Fewer electrons are present in p -type silicon than in undoped silicon because of the recombination of some electrons with the enhanced population of holes.

Junctions

Four different types of junctions can be used to separate the charge carriers in solar cells: (1) a homojunction joins semiconductor materials of the same substance, eg, the homojunction of a p - n silicon solar cell separates two oppositely doped layers of silicon; (2) a heterojunction is formed between two dissimilar semiconductor substances, eg, copper sulfide, Cu_2S , and cadmium sulfide, CdS , in Cu_2S - CdS solar cells; (3) a Schottky junction is formed when a metal and semiconductor material are joined; and (4) in a metal-insulator-semiconductor junction (MIS), a thin insulator layer, generally less than 0.003- μm thick, is sandwiched between a metal and semiconductor material.

Fabrication methods that are generally used to make these junctions are diffusion, ion implantation, chemical vapor deposition (CVD), vacuum deposition, and liquid-phase deposition for homojunctions; CVD, vacuum deposition, and liquid-phase deposition for heterojunctions; and vacuum deposition for Schottky and MIS junctions.

A homojunction can be made on a silicon wafer in the following ways. In the solid-state diffusion process, a *p*-type wafer is heated to a high temperature (ca 850°C) and is placed in contact with the vapor of a desired *n*-type dopant, eg, POCl₃ or PH₃. The dopant atoms thermally migrate or diffuse into the wafer. The dopant concentration converts the top surface layer of the wafer into a *n*-layer, thus forming a *p-n* junction. The *n*-type dopant can be ionized in a vacuum and the ions accelerated to high velocities and injected into the surface of a *p*-type wafer (ion implantation) to form the *p-n* junction. Alternatively, the *p-n* junction can be formed by growing a thin *p*-type layer of silicon on a *n*-type wafer using various techniques. In chemical vapor deposition, a gas that contains a silicon compound and a *p*-dopant is decomposed on the surface of a *n*-type wafer, generally at atmospheric pressure, to form a thin, doped silicon layer. During vacuum deposition, vaporized silicon and *p*-dopant atoms are deposited on the *n*-type wafer in a vacuum. In liquid-phase deposition, the *p*-type layer is grown on the *n*-type wafer from a solution containing silicon and a *p*-dopant. These layers are subsequently heated to permit interdiffusion of the atoms.

Although all of the preceding processes can produce a *p-n* homojunction, the properties required from the junction and the fabrication costs determine the fabrication method. Methods of making other kinds of junctions are similar to those used to make homojunctions, but the materials on either side of such junctions are dissimilar.

Action at the *p-n* Junction. When two pieces of silicon, one *n*-type and the other *p*-type, are brought into contact with one another, initially there are only electron carriers on the *n*-side of the newly formed junction and hole carriers on the *p*-side. This condition causes a large difference in electrostatic charge density across the junction. Immediately after contact, electrons and holes diffuse across the junction in opposite directions. Figure 2a illustrates the distribution of charges in a *p-n* junction solar cell. In the region away from the junction, donor and acceptor ions are neutralized by the presence of free charge carriers. Donor ions are dopant atoms that have contributed electrons to the crystal lattice and thus have positive net charges. Acceptor ions are dopant atoms that have contributed holes to the crystal lattice and consequently have net negative charges. The zone immediately adjacent to the junction is depleted of free charge carriers, which are needed to electrostatically neutralize the ions, and is called the depletion region. The width of the depletion region varies with the concentration of the dopants, but generally is less than 10⁻⁴ cm. In the depletion region, the unneutralized donor and acceptor atoms produce an electric field. This field causes an electrostatic or barrier potential, *V_b*, across the junction. *V_b* is a measure of the potential, which maintains the position of the majority carriers on their respective sides of the junction. Figure 2b–d shows simplified energy band diagrams of the same cell. The uppermost curved line in these drawings denotes the conduction energy level, *E_c*, for the electrons in a solar cell; the lowest curved line denotes the valence energy level, *E_v*, for the electrons. The cell's band gap energy, *E_g*, is represented as the distance between the *E_c* and *E_v* lines. The dotted line indicates the Fermi energy level, *E_f*, ie, the level of energy at which the probability of occupancy of an electron state is 50%; below this level, more than 50% of the available electron states are occupied. The relative horizontal position of the Fermi levels across the junction indicates the presence or absence of induced voltage.

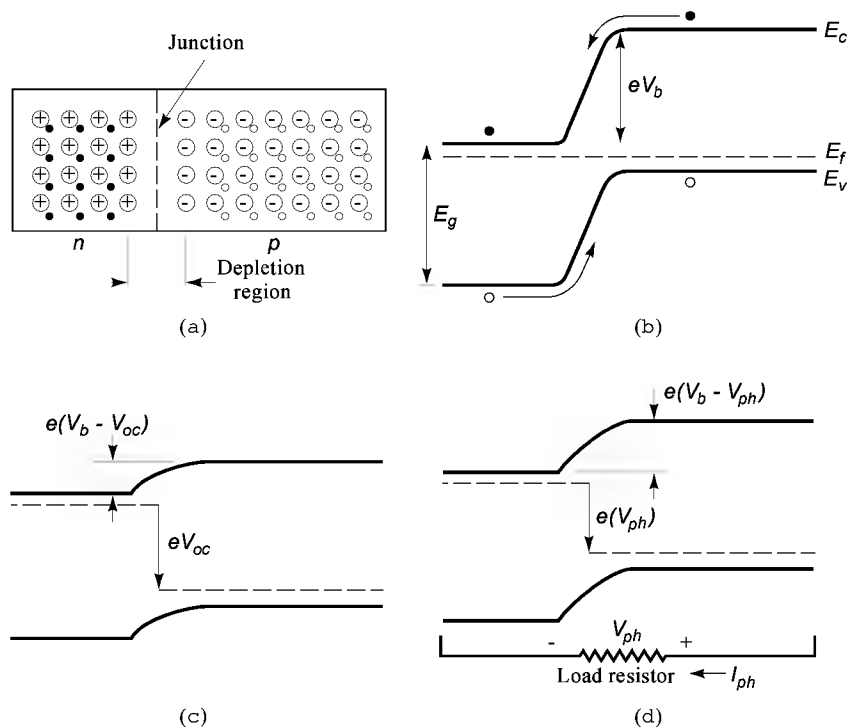


Fig. 2. (a) A schematic diagram of a *p-n* junction, including the charge distribution around the junction, where ⊕ represents the donor ion; ⊖, acceptor ion; •, electron; ◦, hole. (b) A simplified electron energy band diagram for a *p-n* junction cell in the dark and in thermal equilibrium under short-circuit conditions, where *E_f* represents the Fermi energy level. (c) For the same cell in sunlight and under open-circuit conditions where *V_{oc}* is open-circuit voltage. (d) For the same cell in sunlight and under load conditions where *V_{ph}* is photogenerated voltage and *I_{ph}* is photogenerated current.

In the normal operating temperature ranges for solar cells, almost all the extra valence electrons associated with the donor dopant atoms in the *n*-region escape their weak bonds and reach the conduction band, thereby becoming free carriers. In the *p*-region, the holes associated with the acceptor dopant atoms also are free but are in the valence band. The free electrons and holes are the majority carriers in the *n*- and *p*-regions, respectively. The number of majority carriers generally is determined by dopant concentrations and is not sensitive to either thermal or light excitation. However, the number of minority carriers, ie, holes in the *n*-region and electrons in the *p*-region, is sensitive to both thermal and light excitation. This sensitivity affects the operation of the solar cell.

In the dark, all the minority carriers in the *n*- and *p*-regions of a cell are generated by thermal excitation. Once the minority carriers in the *p*-region (free electrons) diffuse through the *p*-region to the depletion zone, they are swept by the built-in barrier potential, *V_b*, across the junction and into the *n*-region. In a similar manner, minority carriers in the *n*-region (holes) are swept into the *p*-region. The energy band diagrams in Figure 2 are for electrons; for holes, the diagrams should be inverted so that holes fall downhill as they are swept from the *n*-region into the *p*-region. This flow of thermally generated minority carriers in a solar cell in darkness is called a dark-drift current. The current is balanced by an opposing flow of charge carriers across the junction, ie, a back-diffusion of majority carriers, which are electrons crossing from *n*-region to *p*-region and holes crossing from *p*-region to *n*-region, flowing against the electrostatic potential *V_b*. Because majority carriers far outnumber minority carriers, *V_b* adjusts itself so that these two opposing currents are equal in magnitude, and equilibrium is established between the *n*- and *p*-region with the Fermi levels lining up on both sides of the junction. In the dark there is no output voltage and no net current.

When sunlight falls on a *p-n* junction solar cell while it is short-circuited, the magnitude of *V_b* remains essentially the same as it was in darkness. Because the diffusion of majority current only varies with *V_b*, the majority current does not change. However, additional minority carriers are formed by

photons, which are absorbed in both the *n*-region and *p*-region, and are swept across the junction. The flow of these minority carriers is in the same direction as the dark-drift current and is a net current flow, ie, the photogenerated short-circuit current, *I*_{sc}.

The same cell in sunlight, but under an open-circuit condition, cannot develop a net current flow. Instead, the cell achieves an equal internal flow of majority and minority carriers across the junction by reduction of its electrostatic potential from its original value *V*_{*b*}, thus there is no net current. This reduction allows a much larger diffusion current, which balances the photogenerated minority current *I*_{sc} (Fig. 2c). The decrease in barrier potential from *V*_{*b*} to *V*_{*b*} − *V*_{*oc*} causes an open-circuit voltage, *V*_{*oc*} of the same magnitude across the open-circuit terminals of the cell. Fermi levels are offset by *V*_{*oc*} or the net-induced voltage. Therefore, *V*_{*oc*} is equivalent to the reduction from the built-in barrier potential *V*_{*b*}, ie, the larger *I*_{sc}, the greater the barrier potential reduction. Although theoretically the maximum value of *V*_{*oc*} is *V*_{*b*}, this condition is possible only in a very high concentration of sunlight, ie, with extremely large *I*_{sc}. Furthermore, because *V*_{*oc*} has a value equal to or less than the barrier potential, cells with a large *V*_{*b*} usually produce a large *V*_{*oc*}, which is why cells with larger band gaps have large values of *V*_{*oc*}.

Under both short-circuit and open-circuit conditions, a solar cell produces no electric power, the power is consumed internally in the cell and is dissipated as heat. When a resistive load is connected to a cell in sunlight, a photogenerated voltage, *V*_{*pb*} is induced across the load and a current *I*_{*pb*} flows through it. The existence of *I*_{*pb*} requires that the flow of majority carriers be reduced from that in the open-circuit condition; there must be a higher barrier potential than in the open-circuit case (Fig. 2d). This higher barrier potential (*V*_{*b*} − *V*_{*ph*}) indicates a smaller reduction from *V*_{*b*}. Since the photogenerated voltage that is induced across the cell is the same as the photogenerated voltage *V*_{*pb*}, the magnitude of *V*_{*pb*} is always less than *V*_{*oc*}. The Fermi levels are offset by *V*_{*pb*}.

The photogenerated current *I*_{*pb*} is in the same direction as *I*_{sc}, but is always less than *I*_{sc} because the barrier potential under load conditions is always less than *V*_{*b*}, which results in a larger flow of majority carriers than that in a short-circuited cell. Thus, when a solar cell is under load, the current and voltage are always less than *I*_{sc} and *V*_{*oc*} respectively; this condition is the curve-factor loss. Depending on the characteristics of the particular *p*–*n* junction and on the cell operating conditions, there is an optimal load resistance that maximizes the power output of the cell, ie, the product of its current and voltage.

When the temperature of a solar cell rises, cell conversion efficiency decreases because the additional thermal energy increases the thermally generated minority (dark-drift) current. This increase in dark-drift current is balanced in the cell by lowering the built-in barrier potential, *V*_{*b*} to boost the majority diffusion current. The drop in *V*_{*b*} causes a decrease in *V*_{*oc*} and *V*_{*pb*}. Therefore, a cell's output, ie, the product of *V*_{*pb*} and *I*_{*pb*} decreases with increasing cell temperature. *I*_{*pb*} is less sensitive to temperature changes than *V*_{*pb*} and actually increases with temperature.

Efficiency

The most efficient silicon cells produced are based on *p*–*n* homojunctions and convert 23.1% of the energy in incident light set to simulate the global air mass (AM) 1.5 spectrum, an artificial reference spectrum used to standardize measurement of PV power, with an intensity of 1000 W m² at 25°C (3). This is the definition of peak sunlight test conditions. In theory, silicon *p*–*n* junction solar cells can convert a maximum approaching 26% of the energy in AM 1.5 sunlight to electricity (4,5). Approximately 75% of the energy in sunlight is lost to factors intrinsic to the silicon material.

The largest single loss is caused by overly energetic sunlight; as much as 32% of the energy in sunlight is lost because sunlight contains many photons with energies larger than *E*_{*g*} for silicon, ie, >1.1 eV. Generally, only one electron–hole pair can be created by a photon regardless if its energy is more than 1.1 eV. The energy greater than *E*_{*g*} is lost in the conversion process and only produces heat within the cell. Another 24% of sunlight is useless for the conversion process in silicon because the photons have energies less than *E*_{*g*} and therefore are unable to create electron–hole pairs.

The maximum power is always less than the product of *I*_{sc} and *V*_{*oc*}. Other losses are caused by effects such as reflection, additional electrical losses, and shading by contact bars and fingers.

In comparison, *p*–*n* homojunction cells made of more costly semiconductor materials, eg, indium phosphide, InP, and gallium arsenide, GaAs, which have energy gaps of 1.2–1.4 eV, and maximum theoretical conversion efficiencies of ca 28–30% (6,7), depending on the device construction and layering of junctions. Materials with even higher band gap energies, eg, copper gallium diselenide, CuGaSe₂ (*E*_{*g*} = 1.7), have higher photovoltages but much lower photocurrents, which produce lower power-conversion efficiency in sunlight. The reduction in photocurrent occurs because there are fewer numbers of solar photons energetic enough to create electron–hole pairs in such materials.

Splitting the Solar Spectrum. Solar cells can be tailored to be optimally efficient in certain limited energy ranges. Because a single solar cell with one junction can convert only a fraction of the incident sunlight into useful energy, dividing the solar spectrum into energy ranges and making each range incident upon appropriately designed cells can result in increased conversion efficiency (8). In one approach, solar cells with different energy gaps are stacked in tandem so that the cell facing the sun has the largest energy gap. The top cell absorbs all the photons at and above its energy gap and transmits less energetic photons to the cells below. The next cell in the stack absorbs all the photons with energies equal to or greater than its energy gap, and transmits the rest downward in the stack, etc. In principle, any number of cells can be used in tandem. The maximum achievable efficiency of a three-cell stack is ca 35–40% at AM 1.5 with a thousandfold concentration of sunlight (9,10).

Designing tandem cells is complex. For example, each cell must transmit efficiently the insufficiently energetic photons so that the contacts on the backs of the upper cells are transparent to these photons and therefore cannot be made of the usual bulk metal layers. Unless the cells in a stack can be fabricated monolithically, ie, together on the same substrate, different external load circuits must be provided for each cell. The thicknesses and band gaps of individual cells in the stack must be adjusted so that the photocurrents in all cells are equal. Such an optimal adjustment is especially difficult because the power in different parts of the solar spectrum varies under ambient conditions. Despite these difficulties, there is potential for improvement in cell conversion efficiency from tandem cells.

Commercial Silicon Solar Cells

Silicon cells are hundreds of micrometers (μm) thick in order to facilitate handling with minimal breakage, although most solar radiation is absorbed in the first 20–30 μm. Light penetration decreases exponentially, proportional to *e*^{−ατ}, where α is the absorption coefficient of a material and τ is its thickness. The values of α for a given material vary with the wavelength of incident radiation; in silicon, α is 10³–10⁵ cm over most of the range of usable solar radiation.

The junction in a silicon cell usually is ca 0.2–0.5 μm from the surface of the cell. This shallowness minimizes the creation of photogenerated carriers in the layer above the junction. Such charge carriers might reach the crystal surface and recombine before being swept through the junction or recombine as a result of the high doping level, which would greatly decrease the photocurrent and therefore the cell efficiency. The crystal surface has many broken bonds that act as recombination centers. The top layer is thin and most of the charge carrier generation occurs below the junction. However, thinning the top layer increases its sheet resistance which is approximately inversely proportional to layer thickness. If the charge carriers that are swept from the lower layer into the top layer must pass through a large resistance before they are collected, much of the energy is dissipated. In conventional silicon cells, a comb or narrow metal grid lattice is connected to a current-carrying bus to collect charge carriers from the side of the cell facing the sun (Fig. 3). The fingers are small enough in total area so that minimal cell area is in their shadow.

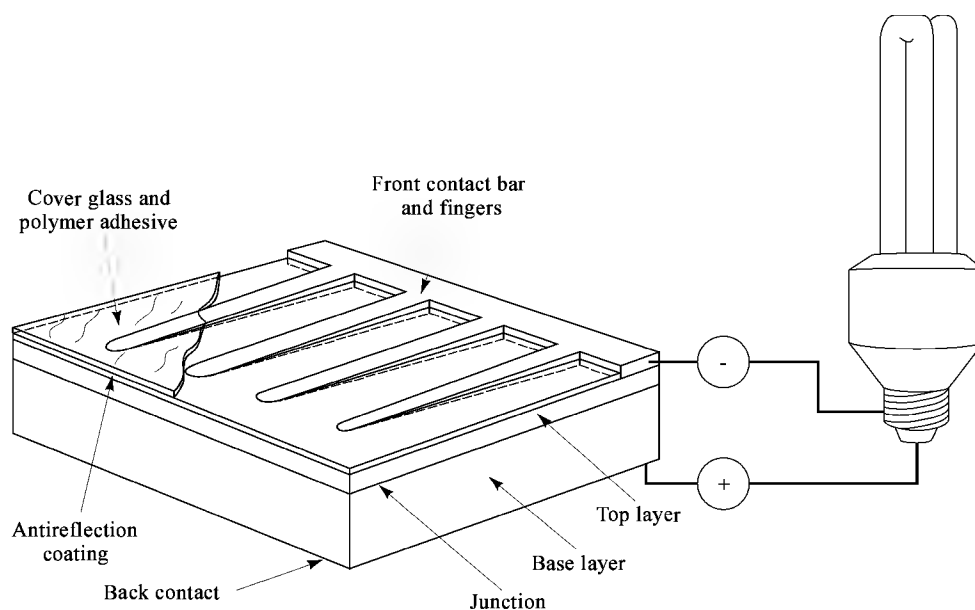


Fig. 3. Basic design of a commercial silicon solar cell.

Antireflection coatings are used over the silicon surface which, without the coating, reflects ca 35% of incident sunlight. A typical coating consists of a single layer of a transparent dielectric material with a refractive index of ca 2, which is between the index of silicon and air or cover material. Materials such as titanium dioxide, TiO_2 , tantalum pentoxide, Ta_2O_5 , or silicon nitride, Si_3N_4 , ca 0.08- μm thick are common. The coating and a physically textured cell surface, such as small shaped facets, can reduce reflectivity to <5% (11). Solar cells are often encapsulated with a transparent adhesive to minimize potential degradation resulting from exposure to corrosive environments. The adhesive also bonds them to a protective cover glass, if one is used. The most widely used encapsulation materials are formulations based on silicone rubber or ethylene-vinyl acetate (EVA), which have proven to be of high durability over periods exceeding 15 years.

MANUFACTURING TECHNOLOGY

In the early 1970s, the first companies to apply low cost, mass production techniques to photovoltaics, a technology that had previously been considered an exotic aerospace technology, emerged. These techniques included the use of electroplated and screen printed metal paste electrical conductors, reflow soldered ribbon interconnects, and by 1977, low cost, automobile windshield-style, laminated module construction. Such processes benefitted from a substantial existing industrial infrastructure, and have become virtually ubiquitous in the present PV industry.

Opportunities for Cost Reduction. Solar cells can be made from several different semiconductor materials, and these materials are available in a variety of physical states: single-crystal, polycrystalline (many small crystals), or amorphous (noncrystalline, eg, glass). Cells are assembled into packaged modules consisting of cells, interconnects, power leads, and a transparent cover or optical concentrator, depending on the type of module. Modules fall into two broad categories: flat-plate modules, used under ordinary sunlight, and concentrating modules, which include lenses to focus sunlight onto the solar cells for locations with generally clear sky conditions.

There is more than one way to make PV systems cost effective, ie, by making more efficient and less expensive devices, by stimulating the market toward higher sales in order to justify production volume increases to achieve economies of scale, and by combinations of these options. In any case, modules must operate reliably for long periods of time.

The cost of a PV device is determined by several factors. These include the kind of materials used and the amount of materials required, choice of substrates, device design, and fabrication processes. Crystalline devices are generally more efficient, but thin-film devices are anticipated to cost less in flat-plate configurations. The use of concentrated light permits retention of efficiency with simultaneous reduction in cost.

TYPES OF SOLAR CELLS

There are three basic technology options for making solar cells with dozens of variations on each. These approaches are conveniently grouped as follows: thick ($\sim 300\ \mu\text{m}$) crystalline materials, concentrator cells, and thin ($\sim 1\ \mu\text{m}$) semiconductor films.

Thick Crystalline Materials.

Silicon. Crystalline silicon technology is the worldwide industry standard. The basic R&D was performed in the 1950s by Bell Laboratories and first commercialized in the 1960s for space power applications. Single-crystal silicon technology cost is challenged by manufacturers of cast block polycrystalline silicon, which is also in the thick category. The casting technique is less capital intensive than single-crystal growth but results in lower efficiency. The world's largest PV manufacturing plants use these crystalline silicon technologies and the fully burdened production cost differential is small. A doubling in sales volume and plant capacity along with investment in automation, wire saws for wafer cutting, and modern material handling methods might reduce cost about 40%, but at that point fundamental raw material costs and packaging (qv) requirements limit further cost reduction. This puts emphasis on greater utilization efficiency with the same essential starting materials and encourages efforts to develop alternative approaches. The options of ribbon silicon technology or concentrator cell approaches, which use thick cells based on silicon or gallium arsenide, are approaches toward circumventing the raw material cost problems.

The total cost of solar cells made from ingots reflects the costs of the silicon raw material used in forming an ingot, cutting and etching thin silicon wafers from the ingot, fabricating and encapsulating the cells, and assembling them into modules. An attractive cost-reducing approach is to grow good quality crystalline sheets directly from molten silicon. Smoothly grown sheets ca 100- μm thick require little or no cutting and polishing and incur little waste. Cells of relatively high efficiency have been fabricated from silicon sheets that are produced by growth-from-melt processes, eg, edge-defined, film-fed growth (EFG), and web-dendritic growth. The EFG technique consists of pulling a ribbon of molten silicon through a slotted die. The ribbon, which generally is polycrystalline with large crystallites, is shaped by the orifice in the die which is made from a material, eg, graphite, that is wetted by molten silicon. In the web-dendritic growth technique, silicon sheet, which generally is single crystalline with a twin plane passing through the center, is grown from the melt in a sheet between two silicon fibers.

Unfortunately, silicon sheets that are made from a melt often develop a granular or semicrystalline texture. The boundaries of the individual crystallite grains have many detrimental effects on photovoltaic conversion, ie, crystalline grain boundaries provide recombination centers for holes and electrons; boundaries behave as barriers to the movement of charge carriers, eg, horizontal grain boundaries isolate some charge carriers from the cell junction, and charge carriers that cross the junction must pass through resistive grain boundaries before reaching a contact sheet or finger; charge carriers can leak across the junction along grain boundaries; and undesirable impurities often diffuse much more rapidly along the grain boundaries than in single crystals, sometimes resulting in short-circuiting of the junction. Such difficulties have slowed the commercialization of these technologies. However, hydrogen passivation of grain boundaries has been effective in minimizing these effects and numerous groups are working to scale-up sheet and ribbon

technologies to mass production ($\text{MW}_{\text{peak}} \text{ yr}$) levels. Figure 4 displays a schematic for these basic alternatives.

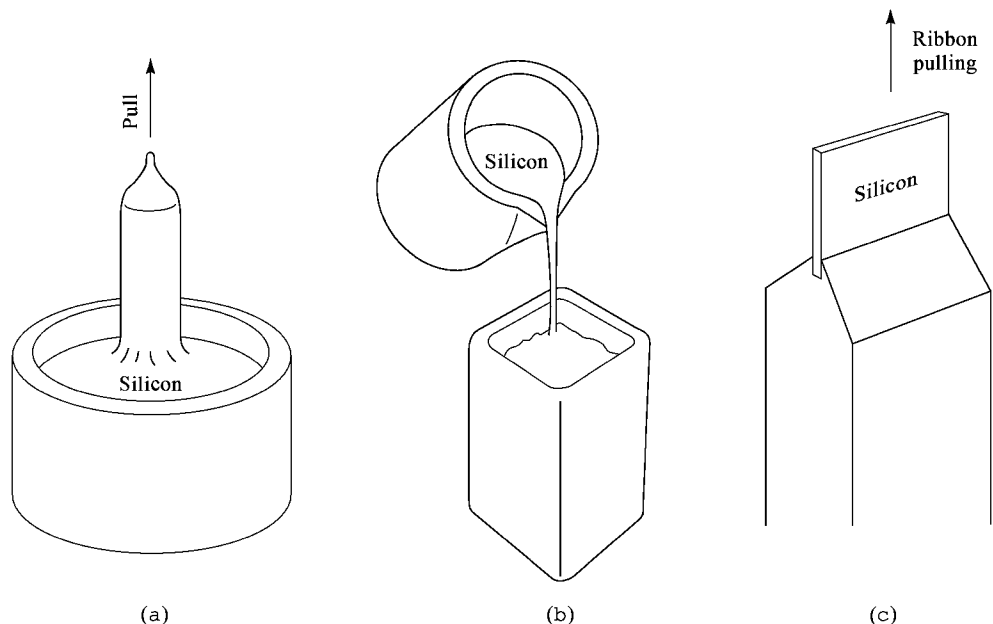


Fig. 4. (a) Single-crystal growth (Czochralski method), (b) casting, and (c) ribbon formation options for silicon.

Crystalline silicon technology is the most mature and best understood of PV technologies. Researchers have identified the principal barriers that limit efficiency and, as a result, since the mid-1980s laboratory cells have climbed from 18 to $\sim 23\%$ and commercial production from 12 to $\sim 15\%$. This is a particularly impressive achievement since crystalline silicon was regarded as mature in the early 1980s.

A concept gaining support is a hybrid approach to making thick crystalline silicon efficient in thin layers. Although conventional crystalline silicon cells have gone from 400–600- μm thick to 200–300- μm , thin-film crystalline silicon cells have reached 10% efficiency while being only 10- μm thick.

Gallium Arsenide. Gallium arsenide is a promising material for gaining the advantages of high efficiency. It is superior to silicon in several respects. The E_g of GaAs, ca 1.4 eV, is higher than that of silicon and is in the range that provides the highest calculated conversion efficiency for a single-junction cell. Because of this high efficiency and the fact that it does not decline as rapidly as that of silicon cells with increasing temperature, GaAs single-crystal cells are attractive for use as concentrator cells.

Gallium arsenide solar cells advanced in the 1980s for space use because they weighed much less than silicon cells of similar output, since GaAs absorbs sunlight much more strongly than silicon. Thinner layers, ca 1–2 μm thick, can be used (12). Although the lower band gap of silicon allows more photocurrent than GaAs, the higher V_{oc} for GaAs cells more than compensates for the decreased photocurrent and results in higher conversion efficiency.

Two different high efficiency types of GaAs cell structures are being developed. These are GaAs homojunctions with a top window layer of $\text{Ga}_{1-x}\text{Al}_x$ and GaAs shallow homojunctions. Each type is being developed in order to minimize the effects of the recombination of charge carriers at the site of broken bonds in the top surface. Such recombination effects, which are more severe in GaAs than in silicon because of the shallower absorption depth of sunlight in GaAs, can severely reduce the photocurrent and therefore the conversion efficiency. Reported efficiencies of GaAs cells are ca 27.5% at AM 1.5 with a sunlight concentration of 205 and cell temperatures of 25°C (3). Silicon cells have achieved conversion efficiencies of 26.5% at AM 1.5 with a 140-fold concentration of sunlight and cell temperatures of ca 25°C. Sophisticated multijunction solar cells such as GaAs–GaSb and GaInP–GaAs, with expected practical limits of about 40%, have exceeded 30% efficiency in laboratory configurations (13).

Concentrator Cells and Systems. Concentrators circumvent the problem of high semiconductor material cost by using mirrors or lenses to concentrate sunlight on small surface areas of more expensive solar cells. Concentration allows more power to be produced from a given amount of photosensitive material. This can solve the cost problem but may create new challenges. First, concentrators only operate efficiently in areas with clear sky conditions (output is ~ 0 on hazy days, unlike flat-plate modules), and large scale ($>500 \text{ W}_{\text{peak}}$) is generally necessary to permit the cost-effective installation of tracking structures. Within certain limits, the concentration strategy can result in cost reduction. The photocurrent generated in a solar cell increases linearly with increasing light intensity within practical limits of concentration, and photovoltage increases roughly as the logarithm of the increase in intensity up to the built-in barrier voltage. The result of these combined effects can be to increase the conversion efficiency as well as boost net electrical output. However, the cells must be specially designed to minimize resistive losses because of the increased photocurrent densities. In addition, the concentration of sunlight often causes increased cell temperatures, so much so that conversion efficiency suffers rather than benefits from the concentration; therefore, some concentrator cells are equipped with cooling systems. Although it may be possible to utilize both the electricity and the heat produced in concentrator cell systems, the technology involved in such utilization has not been well developed.

Concentrator optics vary from low ratio designs, eg, concentration of sunlight of an order of magnitude by Winston collectors, which do not require elaborate tracking of the sun, to much higher ratio systems based on parabolic mirrors or Fresnel lenses and which require precise, two-axis tracking (14). Three types of concentrator systems are being developed which operate at low level (<30 times), mid-level (100–400 times), and high level (>400 times) sunlight concentrations. The cell specifications and engineering requirements for each of these types of systems are quite different. Specially designed silicon has shown potential for use in concentrator systems.

Thin Film. In the thin-film approach, raw material usage is generally more than two orders of magnitude less and patterning is more direct. In some thin-film approaches, certain individual layers may be only 50 atoms thick, which means that large-area uniformity of coating is the key to success. These coatings must be both optically and electrically uniform over areas the size of about a square meter. The technical decisions are complex and may be ordered as follows: (1) What substrate is to be coated? The principal choices are glass, steel, ceramic, or plastic. (2) What materials are to be deposited? The principal semiconductor options are amorphous and polysilicon, cadmium telluride, copper indium diselenide, and alloys of these basic options. The most significant conductor options are silver, nickel, aluminum, tin oxide, zinc oxide, indium oxide, and some alloys of these choices. (3) What deposition process is to be utilized? The options are vacuum evaporation, sputtering, glow discharge, chemical vapor deposition (CVD), electroplating, spraying, and screen printing. (4) How are the layers to be patterned? These options include screen printing, laser scribing, mechanical scribing, and photolithographically defined wet etching.

Good solar cell results have been obtained from cells of materials, including polycrystalline silicon, amorphous silicon–hydrogen ($\alpha\text{-Si:H}$) alloys, Cu_xS –CdS, CuInSe_2 –CdS, and CdTe.

Polycrystalline Silicon. Thin-film polycrystalline silicon cells should be at least 10–30- μm thick for efficient absorption, depending on the effectiveness of the light trapping techniques (15). Such thickness requires photogenerated charge carriers to diffuse similar distances to reach the junction. Minimization of recombination and loss of these carriers can be effected by use of efficient polycrystalline layers, which have as few horizontal grain boundaries as possible within the absorption zone; the preferable condition is that of crystallite grain boundaries that are vertical or normal to the substrate or the junction. The lateral extent of a crystallite should generally be at least three or four times greater than the absorption distance, so that the vertical grain boundaries do not significantly reduce the photocurrent as a result of recombination. The most efficient thin-film polycrystalline silicon cells

produced are ca 15% efficient at AM 1.5, which is lower than the best efficiency for single-crystal silicon but of significance to opening options for cost reduction. These polycrystalline, thin-film silicon cells are formed by CVD of silicon or crystallization from supersaturated silicon solutions onto ceramic or glass.

Amorphous Silicon. Amorphous alloys made of thin films of hydrogenated silicon (α -Si:H) are an alternative to crystalline silicon devices. Amorphous silicon alloy devices have demonstrated small-area laboratory device efficiencies above 13%, but α -Si:H materials exhibit an inherent dynamic effect called the Staebler-Wronski effect in which electron–hole recombination, via photogeneration or junction currents, creates electrically active defects that reduce the light-to-electricity efficiency of α -Si:H devices. Quasi-steady-state efficiencies are typically reached outdoors after a few weeks of exposure as photoinduced defect generation is balanced by thermally activated defect annihilation. Commercial single-junction devices have initial efficiencies of ca 7.5%, photoinduced losses of ca 20 rel %, and stabilized efficiencies of ca 6%. These stabilized efficiencies are approximately half those of commercial crystalline silicon PV modules. In the future, initial module efficiencies up to 12.5% and photoinduced losses of ca 10 rel % are projected, suggesting stabilized module aperture-area efficiencies above 11%.

As with all thin-film PV technologies, the projected manufacturing costs of α -Si:H alloy PV modules fall rapidly with annual manufacturing volume, ie, MW_{peak} yr. The primary driver of this volume cost reduction is the volume–cost relationship of commercially available thin-film processing equipment. Thin-film coating machines often have capacities equivalent to 3–5 MW_{peak} yr, so that manufacturing economies of scale are more fully realized at the 10–15 MW_{peak} yr range. Lower manufacturing output rates generally suffer from high capital equipment costs, due to the large capacity excess to be depreciated, or from high labor costs, due to reliance on manual processing in lieu of expensive large capital equipment.

Amorphous silicon technology has advanced considerably from the first generation of single-junction α -Si:H products introduced in the 1980s. Throughout the 1980s, single-junction α -Si:H modules were at best 5.5% efficient, and often less than 4% efficient in actual field use. Single-junction α -Si:H technology and α -Si:H manufacturing techniques and equipment were successful in yielding low module fabrication costs (\$ m²), but low module efficiencies resulted in high cost per unit power (\$ W_{peak}) generally well above that of crystalline silicon. Also, the first generations of thin-film module encapsulation techniques in many cases did not adequately protect the active layers and severe (15–35 rel %) losses in power occurred with prolonged outdoor exposure. Overall, the first generation of single-junction amorphous silicon failed to achieve high power, low cost PV power modules with consistent high manufacturing production yields. As a result of these experiences, improvements were made in materials processing, optical design, cell interconnection, and module encapsulation, which significantly improved single-junction α -Si:H module performance and are valuable in the design of newer thin-film module products.

The key determinants of future cost competitiveness of α -Si:H PV technology are α -Si:H deposition rates, module production yields, stabilized module efficiencies, production volume, and module design. Reported α -Si:H deposition rates vary by more than a factor of 10, but most researchers report that the high quality films necessary for high stabilized efficiencies require low deposition rates often due to high hydrogen dilution of the Si (and Ge) source gases (see SEMICONDUCTORS, AMORPHOUS).

Copper Sulfide–Cadmium Sulfide. This thin-film solar cell was used in early aerospace experiments dating back to 1955. The Cu_xS band gap is ca 1.2 eV. Various methods of fabricating thin-film solar cells from Cu_xS–CdS materials exist. The most common method is based on a simple process of serially overcoating a metal substrate, eg, copper (16). The substrate first is coated with zinc which serves as an ohmic contact between the copper and a 30- μ m thick, vapor-deposited layer of polycrystalline CdS. A layer is then formed on the CdS base by dipping the unit into hot cuprous chloride, followed by heat-treating it in air. A heterojunction then exists between the CdS and Cu_xS layers.

Most of the sunlight is absorbed by the Cu_xS layer, therefore the CdS layer does not have to be very thick because it does not absorb much sunlight. However, because of the fabrication process, the CdS layer must be thick enough to prevent Cu_xS from diffusing along the grain boundaries deeply enough to reach the metal substrate, thereby short-circuiting the heterojunction.

Laboratory thin-film cells that are fabricated using this cell structure demonstrate a conversion efficiency of slightly greater than 10% (12). Unfortunately, efforts to create a device that is stable for long periods have been unsuccessful and little effort to develop this material is underway.

Cadmium Telluride. Cadmium telluride [1306-25-8], CdTe, is another promising thin film. CdTe is a well-known semiconductor often used in high performance infrared sensors. CdTe absorbs visible light very strongly, and very thin films (1–2 μ m) are sufficient to absorb most sunlight. Small-area thin-film CdTe solar cells have been fabricated with sunlight-to-electricity conversion efficiencies near 16%, comparable to crystalline silicon solar cells in large-scale manufacturing. Large-area monolithic integrated CdTe modules have been fabricated with efficiencies of ca 10%, comparable to crystalline silicon modules commercially available.

CdTe technology is particularly attractive because of its potential for low cost manufacturing techniques, including electrodeposition, spraying, and screen-printing–nonvacuum methods that are less capital-intensive than most other thin-film processes. Given the low manufacturing costs possible with thin-film processing techniques and the high efficiencies of CdTe photojunctions, it is widely anticipated that in full-scale production CdTe PV modules will be competitive in performance, price, and durability.

Companies in the United States, Japan, and Europe are actively working to commercialize CdTe PV technology. CdTe PV modules are available in limited quantities and sizes from one Japanese manufacturer, and two U.S. companies (Golden Photon in Colorado and Solar Cells Inc. in Ohio) have announced their intention to market CdTe PV products in the near future.

Copper Indium Diselenide. CuInSe₂ (CIS) has proven to be one of the most promising thin-film photovoltaic materials. CIS alloy materials have yielded small-area (ca 1 cm²) laboratory devices with efficiencies in excess of 17% and large-area (ca 0.5 m²) monolithic integrated modules with efficiencies in excess of 11%, and have shown excellent radiation hardness.

CuInSe₂ is an ideal thin-film photovoltaic absorber material. The CIS band gap is near the optimum for photovoltaics, ie, low enough to absorb much of the useful solar spectrum and high enough to generate high operating photovoltages at normal outdoor operating conditions. CIS and its alloys exhibit strong optical absorption; films of <1- μ m thickness are sufficient to absorb most sunlight. The grain boundaries and surfaces of CIS and its alloys are electronically benign, so that simple polycrystalline films yield high efficiency photovoltaic devices without complex grain boundary passivation processing. The defect chemistry of CIS materials is such that its microstructural and compositional tolerance are high, thus compositional variations, eg, Cu–In atomic ratio, of ± 5 atom % can be tolerated. This tolerance of microstructural and compositional variations relaxes the requirements for processing precision and uniformity without significantly impacting the overall device efficiency. High efficiency CIS junctions can be fabricated with relative ease. CIS forms stable, high conductance ohmic contacts with molybdenum metal, and high quality *p–n* junctions can be formed with thin CdS layers and standard transparent conductors like zinc oxide and indium–tin oxide. CIS PV junctions are stable. CIS devices can be made with simple monolithic interconnects between individual cells, allowing the fabrication of large monolithic integrated modules that minimize manufacturing labor requirements. CIS modules can be packaged simply using the pottant–glass encapsulation techniques proven for crystalline silicon PV products. Packaged CIS modules have proven durable in long-term outdoor tests (see PACKAGING, ELECTRONIC MATERIALS).

The technology forefront for CIS photovoltaics is twofold: graded alloys and large areas. CIS PV technology is evolving toward graded CIS alloy materials containing sulfur and gallium. CIS alloy absorbers can yield significantly higher photovoltages at slightly lower photocurrents, thus providing higher overall device efficiencies. Sophisticated composition grading allows the introduction of back surface fields and front surface minority carrier recombination control to yield higher efficiencies. These graded alloy materials can be formed by the co-evaporation techniques in use for laboratory experimentation, but are difficult to replicate on large areas using industrial physical vapor deposition processes that might be used in manufacturing.

CIS PV technology is evolving toward larger areas as the technology progresses from research laboratories to production factories. Processing on large-area substrates and production of large-area modules are essential in achieving low manufacturing costs. The point–source co-evaporation laboratory techniques used to develop the graded CIS alloys necessary to achieve the highest efficiencies are difficult to control on large areas. Physical vapor deposition (PVD) techniques have been developed to deposit basic CIS materials on large areas, but these PVD techniques are not well suited to the deposition of graded alloy materials. The PVD techniques generally require multistep processing and high substrate temperatures. Multiple steps complicate

the fabrication sequence and add cost. The high substrate temperatures required for optimal PVD materials quality limit the options for usable substrates. Thus, CIS is a promising photovoltaic material, but improved processing techniques are needed to achieve commercial production of advanced high efficiency CIS alloy materials. Table 1 summarizes the laboratory and commercial status of significant PV technology.

Table 1. Photovoltaic Technology Overview

	Flat plate										Wafer concentrator	
	Thick film							Thin film			Single-jun- ction Silicon	Multi-j unction III–V
	Self-supported				Substrate			Alloy		Compound		
type	Cz silicon	Poly-Xtal Si	EFG	Web	MIS Si	Si film	Spheral Si	Amorphous	CIS	CdTe		
initial PV demon-stratio n	1954	1976	1974	1959	1981	1987	1983	1976	1975	1971	1954	1965
theoretical efficiency, ^a %	~26	~22	~20	~24	~24	~20	~18	~20	~22	~25	~26	~40
state-of-art cell, %	23.1	17.7	~16	18.3	~16	~15	~12	13	16.4	15.8	25.6 at 140 $\times$	32.6 at 100 $\times$
state-of-art module, %	18.2	12.2	~11.5	~15	na	~11	~10	~10	~11	~10	20.3 at 80 $\times$	~18 at 100 $\times$
commercial module, %	12.5	11	10.5	na	10.5	10.5	na	5	na	na	17 at 30 $\times$	na
representative industrial participants	BHEL	DASA		Ebar a	Nuke m	Astropw r			ISET	BP Solar	Entech	na
	BP Solar	Eurosolare	ASE Am-eric a					Advanced PV Systems	EPV		SEA Corp.	
	CEL	Kyocera					Texas Instru- ments	Fuji Elect.	Siemens	Golden Photon	Sun Power	
		Photowatt						Intersolar	SEO Solar	Matsushita		
	Heliodi- namica	Shell R&S						Kanekafuchi				
	Helios	Solarex						NAPS France	Lockheed Martin	NAPS Finland		
	Isofoton							Sanyo		Solar Cells		
	Sharp							Solarex				
	Siemens							Taiyo Yuden				
	Solec							USSC				
								ECD				

^a Air mass 1.5, 1000 W m², at 25°C.

ELECTROCHEMICAL PHOTOVOLTAIC CELLS

The application of photoelectrochemistry in solar energy conversion technologies includes biomass conversion, photoelectrolysis, photogalvanic cells, electrochemical photovoltaic cells, etc (17–19). In electrochemical photovoltaic cells, electric energy is converted directly from sunlight by absorption of light in a semiconductor electrode. In many respects, these cells closely resemble conventional solid-state cells, except that the charge-separating barrier layer is formed at the interface between a semiconductor surface with a liquid electrolyte. When sunlight is incident on the semiconductor electrode, free holes and electrons are created. The relevant minority carriers must migrate to the interface and be separated; these carriers then react with the electrolyte either through oxidation or reduction. The counterelectrode reverses the reaction, thereby maintaining the electrolyte balance. The semiconductor electrode material may be either polycrystalline or amorphous material because in some cases the poorer material properties cause relatively little degradation of conversion efficiencies. In addition, incorporation of a third electrode may make possible *in situ* storage. The main disadvantage of these cells is the instability of the semiconductor electrode, especially under sunlight, for extended periods of operation. Electrochemical cells could be inexpensive, since the electrode–electrolyte barriers usually are easy to form, but appropriate deployment strategies have not yet been identified. In addition, the stability problems encountered to date (ca 1995) have been extensive.

BALANCE OF SYSTEMS

A solar photovoltaic system contains, in addition to solar cells and module(s), an array structure to support the modules, power-conditioning circuitry for control and modification of the output, and a means of storing energy if required. All elements beyond the module are referred to as balance-of-system (BOS) components. The cost of BOS items is nominally about equal to the cost of the PV module. However, the BOS fractional cost contribution can vary from one- to two-thirds of the total installed cost of a system, depending on application. The direct current (dc) produced by solar cells must be changed to alternating current (ac), when the power is to be connected to commonly found utility power loads (20,21). For many applications, electric storage, eg, in batteries, is needed to provide power when the cells are not illuminated, to supplement output of the cells during transient loading, or to aid in utility power system load leveling (22). In addition, a means of regulating the flow of power between the solar array, energy storage, and load is often required (23,24). The development of less expensive storage for the solar photovoltaic system, eg, flywheels, pumped water, and fuel cells (qv), has been described (25–29). The exact mode of storage depends on the particular application of the photovoltaic system. For example, if the photovoltaic system is used in a rural residence, a small automotive battery is often used. However, if the system is associated with a utility grid, the storage is centralized to the main utility where pumped water may be the storage technique.

Material Availability and Environmental Impact

Photovoltaic systems must satisfy four principal requirements before solar photovoltaic conversion can provide a significant portion of general energy needs. The system costs must be low enough to be competitive with other means of energy generation, the amount of energy generated during the life cycle of a photovoltaic system must be substantially greater than the energy required to fabricate the system to meet the criteria of a sustainable technology, the materials used in the cells must be available to generate a substantial portion, ie, at least a few percent of world energy needs, and the fabrication and

utilization of the conversion systems should not cause more environmental problems than other competing energy systems.

Silicon is the second most abundant element in the world and is not toxic. Inherent in the use of materials other than silicon for solar cells are challenges of material availability and environmental safety. In terms of production of CdS-based cells, sulfur is abundant, but the world's resources of cadmium, tellurium, selenium, and indium are much less than those of silicon (30). However, these resources are several orders of magnitude greater than the amount needed to provide photovoltaic power production of 50,000 MW yr. Similarly, although arsenic is plentiful, the supply of gallium for GaAs cells is limited. However, studies have concluded that the gallium supply also is sufficient for substantial manufacturing scale (31).

Although photovoltaic conversion is nonpolluting, environmental, health, and safety aspects must be considered, especially with regard to harmful emission and waste products resulting from the production of the solar cell modules. These considerations apply to silicon cell production, since various processing steps, eg, doping, may involve toxic materials. For materials, eg, CdTe, GaAs, and CuInSe₂, the presence of toxic Cd or As in the material and device processing must be carefully examined. In addition, the large deployment of such cell modules in terrestrial applications occur only when the long-term safety to the environment is assured. It has been shown that, with proper encapsulation and a proactive recycling program, it should be possible to minimize environmental concerns.

Photovoltaic Markets

The character of the PV market has undergone a substantial shift since the 1970s. In 1975, PV systems were installed exclusively to power remote industrial loads such as telecommunications repeaters, offshore aids to navigation, and cathodic protection systems. In the early 1980s, the installation of several multimegawatt centralized utility PV power plants raised expectations of rapid penetration into the potentially large utility market. Sharp reductions in oil prices curtailed most of the PV utility activity, and the balance of the decade saw the development of the off-grid, rural consumer market segment that dominates the PV business.

In the mid-1990s, utility applications have once again begun receiving a great deal of attention due to a profound paradigm shift that appears to be taking place in the utility industry. Rather than replacing or adding large central fossil-fueled or nuclear generation facilities, small PV systems deployed at the outer extremities of the grid can be cost-effectively used to manage demand profiles, defer transmission hardware upgrades, and support electrical service quality (voltage, power factor, etc) during periods of peak demand in locations where the utility grid transmission is unidirectional (32,33).

PV Market Segment Categories. Solar modules are used to provide power to a broad range of industrial, commercial, and consumer systems and products. Most participants in the PV industry use the following categories to describe the various market segments, which group applications by functional product requirement, system type, sales channel, and client base. These include the following: specialties, eg, spacecraft circuits, calculator chips, automobile sunroofs, and building facades; industrial power, ie, telecommunications, warning signal lights, and remote data gathering; rural and off-grid electrification, eg, lighting, water pumping and purification, refrigeration, and recreational travel and boating; consumer convenience, eg, garden and security lighting and small battery charging; and grid-connected power, ie, distributed grid support and peaking power augmentation.

As PV module selling price has decreased, new markets have progressively emerged. This combination of market growth and price decrease has provided a smooth evolutionary path for the commercial industry of the 1990s.

Specialties. The specialty items business segment is principally a collection of niche market opportunities, including aerospace, eg, satellite power systems, and consumer electronics, eg, calculators. The aerospace industry is not much larger in terms of industry-wide revenues in the 1990s than it was in the 1970s. There are a few small specialty firms that produce cells with exacting quality assurance procedures that do not permit significant deviation from the manufacturing processes in use since the 1960s.

Perhaps the most familiar example in the specialty items category is the consumer electronics market which consists primarily of solar-powered calculators and watches. Although volumes are large in terms of units sold, the revenues are relatively small. Further, the competition is fierce for any photovoltaics manufacturer who seeks to sell commodity solar cells to the consumer goods producer.

Industrial Power. The original use for terrestrial solar modules was in industrial applications. These systems employ an array of modules typically between a few hundred watts and a few kilowatts to charge storage batteries, which are then used to power the load. Usually the batteries contain enough reserve to deliver power for up to a week in the case of bad weather. Industrial systems are located in areas where no power grid exists, and where the alternative forms of remote power generation, such as diesel generator sets or thermoelectric generators, are either impractical or economically unfeasible. Desirable module attributes are high efficiency to reduce the size and therefore the cost of the site and mounting structure, and high reliability (34). In general, the cost of the solar array is a minor percentage of the entire project cost, thus the market size is relatively inelastic to module price.

Examples of remote industrial applications are microwave, radio, and cellular telephone repeaters; supervision, actuation, and data telemetry systems for highway warning signs, wellheads, oil and gas pipeline valves, reservoirs, dams, etc; aids to navigation on offshore buoys, lighthouses, oil and gas platforms, bridges, tanks, masts, and railroad crossings; and cathodic protection systems on pipelines (qv), tanks, and wellheads to inhibit corrosion.

Rural and Off-Grid Electrification. Rural consumer applications relate more to people as contrasted with electrical equipment, and in this sense the service provided by the solar system is no longer simply electrical power supply, but rather light, television, drinking water, etc. Because these services require a much broader spectrum of system hardware, eg, pumps, light bulbs and fixtures, or television sets, than industrial systems, and because the local conditions in the various end use regions are so varied, such systems have been most successfully marketed through in-country distributors. These distributors, or system integrators, buy modules from the PV module manufacturers, but design and source the balance of the components by themselves to fit the local requirements. Local content is mandatory for practical as well as political reasons (35,36). The attention placed on ecological and job issues around the world makes it inevitable that government emphasis on local content, including local manufacturing, will increase. The small modular nature of photovoltaic products is a good match to this requirement.

The rural consumer customer base is large and remarkably diffuse. Unlike industrial systems, the price of the solar array comprises a large fraction of the total system price, and is thus highly leveraging to this market segment. The alternative to a PV system, in many cases, is doing without the service or accepting an inferior level of comfort, eg, contaminated water or kerosene lighting. Reducing the up-front transaction cost is important (37).

Examples of rural consumer applications include small (~50–1000 W_{peak}) homelectric or solar home systems including a few fluorescent lights, a television set, and or radio for remote homes and vacation cabins; larger (>10–20 kW_{peak}) systems for a village power supply or to provide power to a field hospital, clinic, or school; water pumping systems for drinking water or livestock watering as well as for low head microirrigation systems; and battery charging systems mounted to the roof or deck of recreational vehicles and pleasure boats, providing power for lights and, perhaps, a television (38–40). The rural market segment is driving the majority of growth in the PV industry. The world bank has estimated that over 1.7 billion people live without any electric service and that the number is increasing. PV production is expanding to meet this demand. Such progress permits expansion of manufacturing capacity and the cost reduction which accompanies both increased production scale and normal learning curve efficiencies. If just one-half of the homes without any electricity were to have 50 W of PV power (10,000 MW), it is estimated that the volume of PV production would be enough to see module prices fall to a level near \$1.00 W_{peak}⁻¹, which translates to an electricity price of approximately \$0.10 kWh⁻¹ ± 25%, depending on the cost of money and BOS requirements.

Consumer Convenience. The solar powered garden light, with its simplicity and design for functionality, was introduced through mass merchandise retail outlets in the mid-1980s. These units helped popularize PV and evolved into broad use for decorative lighting applications and as high performance security lighting systems. These small systems have served as the test for lighting products of all types.

Grid-Connected Power. This group of applications represents the long-term goal of PV technology, to generate large amounts of power in conjunction with the existing utility grid, thereby displacing fossil fuel use. Implementation concepts range from distributed systems, whereby each house is equipped with a solar array of several peak kilowatts, all feeding the grid in parallel, to large centralized stations similar to the multimegawatt stations constructed in California during the early 1980s (41).

Utilities have become familiar with PV in traditional industrial applications such as tower lighting, telemetry, valve and gate actuation, cathodic protection, communications systems, and security lighting. These systems generally include a fixed tilt crystalline silicon array, charge controller, and battery storage, and are individually designed and installed by a network of system integrators. Such applications are cost effective for module prices below \$7–9 W_{peak}⁻¹, and the market is relatively inelastic to pricing (42).

Stand-Alone Microutility Projects. Utilities have also become familiar with PV in a separate group of associated applications which are small ($\sim 1\text{ kW}_{\text{peak}}$) utility-owned stand-alone domestic power supply systems commonly referred to as "service without the wire." Typically, these systems are essentially identical to a remote industrial system, with the addition of an inverter used to power common appliances that require alternating current. Several utilities are investigating the possibility of installing, owning, and operating such systems as a cost-effective alternative to power line extension.

Microutility systems are categorized on the basis of stand-alone character, ie, primary energy supply, as opposed to line-tied or peaking power supply. These systems are installed in locations without any existing primary source of electricity, in some cases as a replacement or in conjunction with small ($<50\text{ kW}$) diesel generator sets. Such PV microutilities are the most cost effective for electrifying isolated islands and villages in developing nations seeking to prevent mass migration of the rural population to overcrowded urban centers. A typical system configuration includes a tracking PV array that charges a storage battery, which in turn feeds a small a-c distribution grid through the use of a static (battery-tied) inverter.

Battery storage is used to furnish energy at night and during short periods of inclement weather. Rather than designing the battery capacity for an extended period of load autonomy, a small backup generator is included to recharge the battery when necessary. This design philosophy results in emphasizing the value of a relatively small battery which charges and discharges a significant portion of its total stored energy every day. The static inverters used in such systems are standard equipment in widespread use for uninterruptible power supplies (UPS). Some additional inverter output protection, eg, surge suppression, measures are generally included by the PV equipment suppliers.

Microutilities are often sited in areas without existing sources of electric power, and often in countries with limited ability to pay the large up-front capital costs inherent to PV systems. Large-scale usage of PV technology is constrained by the lack of sufficient financing needed to cover the high capital cost, which in many cases must be weighed against subjective and abstract societal and political values. As a further complication, rapid government turnover in many developing countries tends to lead to "quick fixes" with low capital costs, rather than solutions which might provide lower life cycle costs realized over a longer period (43).

The key to this microutility market segment, therefore, is financial in addition to technological. It is likely that any significant success in this market will result directly from the creation of a suitable financing strategy.

Grid Voltage Support. This application has been the focus of increasing discussion during the 1990s, precipitated by numerous utility studies detailing a system planning paradigm shift and concern for the consequences of deregulation in the United States. Specifically, the thinking is based on a detailed evaluation of various low capacity (ie, $<1\text{ MW}$) decentralized generation and distribution options, compared to the previous evaluations based entirely on $200\text{--}1000\text{ MW}$ centralized generation increments. This finely tuned evaluation quantifies some of the commonly accepted but previously subjective benefits of PV systems such as modularity, quick construction time, and inherent dispatchability, ie, a local match between PV system daily output profile and summer peaking utility demand contour. Similarly, the relatively high costs associated with regular (eg, noon–7 pm, July–Sept.) overload of end-of-the-grid feeder circuits become visible when analyzed in greater detail. This type of application was first demonstrated in 1982 with a 1 MW installation in the high desert of Southern California next to the Southern California Edison Lugo substation.

In the grid-connected market, energy prices and regulatory policies have a primary impact on the growth rate and form of market penetration. The market is geographically fragmented based on competing energy prices and options for electricity supply as well as the rates of localized demand growth and sunlight availability. In general, U.S. utilities anticipate that photovoltaic systems can compete into the twenty-first century for distributed voltage support generation additions, as PV power system selling prices reach the $\$3.00\text{--}\$6.00\text{ watt}_{\text{ac}}$ total installed system price. In 1994 prices as low as $\$6.25\text{ watt}_{\text{ac}}$ were quoted to the Sacramento Municipal Utility District for its pioneering PV program. The core of the government role in accelerating utilization of PV power is to provide support, especially with regulatory groups to permit early market entry.

The new utility planning model ascribes a significantly higher value to PV-generated electricity than has been previously considered in the context of large-scale utility applications. Furthermore, a study by Pacific Gas & Electric indicated that roughly 10% of the total 2000 feeder circuits could cost-effectively utilize PV systems of this type, ie, $500\text{ kW}_{\text{peak}}$ (32). Therefore, the addressable market only within the PG&E network (constituting $\sim 3\%$ of total U.S. generating capacity) is on the order of $100\text{ MW}_{\text{peak}}$. This market segment exhibits high price demand elasticity, since the initial market penetration occurs on the highest cost, lowest volume segments of the grid.

Distributed Rooftop. Distributed rooftop-mounted PV systems are usually thought of as an alternative to centrally sited PV power plants such as those previously discussed. Each system, based on a flat-plate integrated panel ranging from $1\text{--}3\text{ kW}_{\text{peak}}$, is connected to the utility grid by means of an individual line-tied inverter. As with the grid support application, power from the aggregate arrays serves to decrease the peak summertime demand on the utility central generation capacity. The arguments for this approach can be divided into two groups, namely quantitative and qualitative (44).

Quantitative arguments for rooftop-mounted arrays deal primarily with the cost savings captured by avoiding the need for an array support structure, since the price of the roof has already been paid. Counter-arguments include the lack of economies of scale (as with central systems), costs associated with individual engineering, permitting, and installation of each system, and the fact that roof-mounted systems can be fixed; as opposed to tracking, resulting in a lower capacity factor per installed peak kilowatt. In some areas, such as Japan, population density and the high cost of farmland virtually preclude the use of centralized systems, and strongly favor the distributed rooftop-mounted approach.

Qualitative arguments deal primarily with the sense of ownership and security which result from individually owned generation systems. Additional complexity will arise from the aesthetic criteria specific to both individual homes and the surrounding community. Of course, the autonomy inherent in distributed rooftop arrays probably constitutes an institutional barrier to their acceptance by some utility companies, unless they are involved in the financing and or marketing chain.

Demonstration Projects. Demonstration projects, many government funded, have been important in the development of the solar industry. Throughout the latter part of the 1970s and the early 1980s, there was a concerted effort on the part of the U.S. government to stimulate the growth of the industry through large block purchases of modules. These modules were used for many different systems, most notably to power houses, ie, lighting, refrigerators, and water pumps on various Native American reservations. Substantial solar tax credits covering up to 50% of the initial system cost in some cases also provided a demand side stimulation of the market. Thousands of rural solar home systems were installed during this period. The result of this activity is a solid, long-term performance history of proven PV system performance which has been independently verified by a wide variety of customers. Some forms of accelerated depreciation and long-term power-purchase contracts are in place that deliver equivalent financial incentives, without creating a dependency on one-of-a-kind subsidies.

In Europe and Japan, government-funded demonstration projects are being actively pursued as a method of direct stimulation of the PV industry.

Market Size and Distribution. The market for PV modules has grown at an average annual rate of more than 15% since the 1980s to approximately $68\text{ MW}_{\text{peak}}\text{ yr}$ (1994) (45). With increased involvement of utilities and lending institutions, such as the World Bank, this rate is expected to continue. Figure 5 provides an overview of the historical and projected market growth.

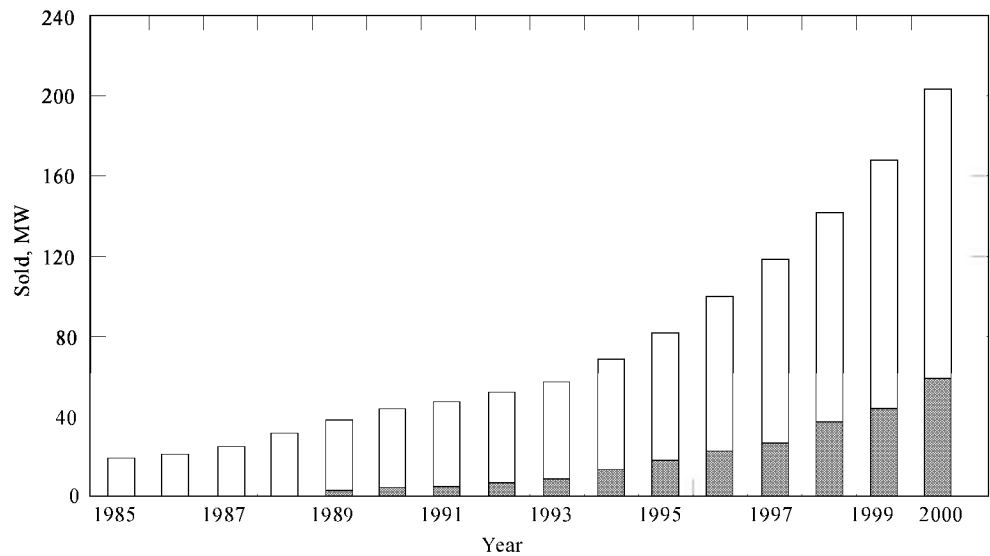


Fig. 5. Worldwide PV business from 1985 through 2000, where (□) represents developing countries and (■), developed countries.

A significant factor in the growth of the industry is an increased participation by local entities in all aspects of the technology from market development to manufacturing, as is represented in the differentiation shown in Figure 5 between products manufactured in developed countries and developing countries (46). This is a reflection of the strong differential growth rate in the rural consumer segment, which accounts for nearly half of the total PV market, versus slightly more than 25% at the start of the 1990s. The rural market segment and the industrial market are expected to remain a significant portion of the total business through the remainder of the twentieth century (Fig. 6).

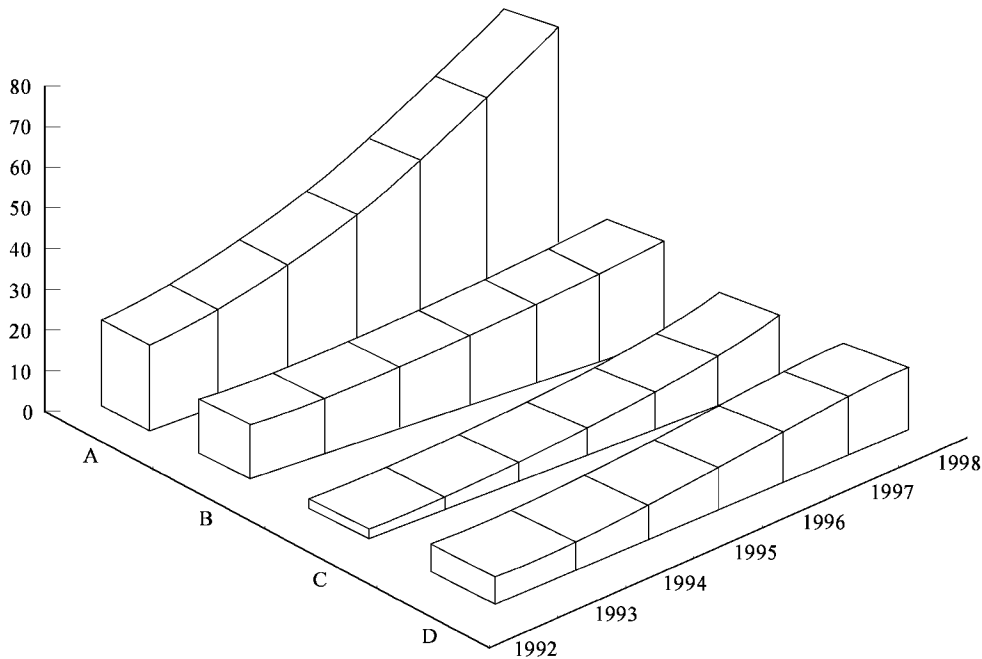


Fig. 6. Past and projected market segment trends (1992–1998), in megawatts, where A represents rural applications; B, industrial; C, utility grid-tie; and D, consumer special.

BIBLIOGRAPHY

"Photoelectric Cells" in *ECT* 1st ed., Vol. 10, pp. 519–542, by V. K. Zworykin and E. G. Ramberg, Radio Corp. of America; "Phototubes and Photocells" in *ECT* 2nd ed., Vol. 15, pp. 396–443, by C. P. Hadley and co-workers, Radio Corp. of America; "Photovoltaic Cells" in *ECT* 3rd ed., Vol. 17, pp. 709–732, by J. C. C. Fan, Massachusetts Institute of Technology.

1. W. Palz, J. E. Rannels, and K. Toma, *12th European Photovoltaic Solar Energy Conference Proceedings*, Amsterdam, the Netherlands, Apr. 11–15, 1994.
2. M. P. Thekaekara, "Survey of Quantitative Data on the Solar Energy and its Spectral Distribution," *Conference of COMPLES*, Dahrn, Saudi Arabia, Nov. 1975.
3. M. A. Green and K. Emery, *Prog. Photovol.: Res. Appl.* **1**, 25–29 (1993).
4. M. Wolf, *Energy Convers.* **11**, 63 (1971).
5. A. Ricaud in Ref. 1.
6. J. J. Wysocki and P. Rappaport, *J. Appl. Phys.* **31**, 571 (1960).
7. *Chem Week*, **54** (Apr. 15, 1981).
8. D. Trivich and P. A. Flinn, in F. Daniels and J. A. Duffie, eds., *Solar Energy Research*, University of Wisconsin Press, 1955, p. 143.
9. J. J. Hovel, *IBM J. Res. Dev.* **22**, 112 (1978).
10. A. Bennett and L. C. Olsen, in Ref. 3, p. 868.
11. R. A. Arndt, J. F. Allison, J. G. Haynos, and A. Muelenllerg, Jr., *Conference Record of the 11th IEEE Photovoltaic Specialists Conference*, Scottsdale, Ariz., May 1975, p. 40.
12. H. Hovel, *Solar Cells*, Vol. 2, Academic Press, Inc., New York, 1975, p. 102.
13. M. A. Green and K. Emery, *Prog. Photovol.: Res. Appl.* **2**, 231–234 (1994).

14. R. W. Winston, *Sol. Energy* **16**, 89 (1974).

15. T. Tiedje, E. Yablonovitch, G. D. Cody, and B. G. Brooks, *IEEE Trans. ED-31*, 711 (1984).

16. A. Rothwarf and K. W. Boer, *Prog. Solid State Chem.* **10**, 71 (1975).

17. J. M. Radovich, *Proceedings of Control Power Systems Conference and Exposition*, Oklahoma, Mar. 1980, p. 114.

18. H. Gerischer, in J. R. Bolton, ed., *Solar Power and Fuels*, Academic Press, Inc., New York, 1976, p. 77.

19. N. N. Lichtin, in *Rural Power Utilities and the Role of Photovoltaics*, South Pacific Institute for Renewable Energy, Tahiti, 1991, p. 119.

20. R. L. Pickrell, G. O'Sullivan, and W. C. Merrill, *Conference Record of the 13th IEEE Photovoltaic Specialists Conference*, Washington, D.C., 1978, p. 984.

21. E. E. Landsman, in Ref. 3, p. 992.

22. S. L. Leonard, *Conference Record of the 12th IEEE Photovoltaic Specialists Conference*, Baton Rouge, La., Nov. 1976, p. 641.

23. E. F. Lyon, L. L. Bucciarelli, and A. E. Benoit, in Ref. 3, p. 1268.

24. D. J. Roesler, in Ref. 3, p. 978.

25. J. O'M. Bocheris, *Energy: The Solar-Hydrogen Alternative*, Architectural Press, London, 1976.

26. D. L. Pulfrey, *Photovoltaic Power Generation*, Van Nostrand Reinhold Co., New York, 1978.

27. J. M. King, Jr., *Proceedings of the 10th Intersociety Energy Conference*, 1975, p. 237.

28. A. R. Millner and T. Dinwoodie, *Conference Record of the 14th IEEE Photovoltaic Specialists Conference*, San Diego, Calif., Jan. 1980, p. 1018.

29. *Energy Storage*, U.S. Dept. of Commerce Document No. Conf. 760212, Executive Summary of Engineering Foundation Conference, Asilomar, Washington, D.C., Feb. 1976.

30. H. Hovel, in Ref. 12, p. 222.

31. *Fourth Annual Photovoltaic Advanced Research and Development Conference Abstracts*, SERI CP11-956, Nov. 1980, p. 245.

32. D. S. Shugar, *5th International Photovoltaic Science and Engineering Conference*, Kyoto, Japan, 1990, p. 871.

33. J. E. Bigger, *Solar Ind. J.*, 27 (1992).

34. *Maintenance and Operations of Stand-Alone Photovoltaic Systems*, Naval Facilities Engineering Command, Southern Division, rev. 1991.

35. M. Hankins, *Solar Rural Electrification in the Developing World*, Solar Electric Light Fund, 1993.

36. N. Lenssen, *Empowering Development: The New Energy Equation*, Worldwatch Paper 111, Nov. 1992.

37. D. Lovejoy, "Electrification of Rural Areas by Solar PV," *International Workshop on Mass Production of Photovoltaics: Commercialization and Policy Options*, Sao Paulo, Brazil, Sept. 16–20, 1991.

38. H. Rodriguez, *Prospects for Photovoltaics Commercialization, Mass Production and Application for Development*, Advanced Technology Assessment System 8, United Nations, New York, 1992, pp. 1–8.

39. B. Chabot, *Rural Electrification Guidebook for Asia and the Pacific*, UN-ESCAP, Bangkok, 1992.

40. H. A. Wade, in Ref. 19.

41. R. J. Arnault, E. Berman, C. Gay, R. E. L. Tolbert, and J. W. Yerkes, "The ASI One-Megawatt Photovoltaic Power Plant," *International Solar Energy Society World Congress*, Perth, Western Australia, Aug. 14–19, 1993.

42. K. Smith, *Survey of U.S. Line-Connected Photovoltaic Systems*, EPRI GS-6306, Palo Alto, Calif., 1989.

43. S. Awerbuch, *Prog. Photovol.* **1**, 153 (1993).

44. J. Day and R. O. Johnson, *Distributed PV Applications*, Report PM-36, Strategies Unlimited, Mountain View, Calif., 1992.

45. C. F. Gay and C. Eberspacher, *Prog. Photovol.: Res. Appl.* **2**, 249–255 (1994).

46. R. H. Annan, W. L. Wallace, T. Surek, E. Boes, and L. O. Herwig, *Department of Energy Review of the U.S. Photovoltaic Industry*, Report ST-211-3488, Solar Energy Research Institute, Golden, Colo., 1989.

General Reference

G. D. Cody and T. Tiedje, in B. Abeles, A. Jacobson, and P. Sheng, eds., *Energy and the Environment*, World Scientific, Teaneck, N.J., 1994.

Charles F. Gay
National Renewable Energy Laboratory

Chris Eberspacher
UNISUN

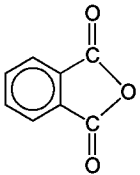
PHTHALIC ACIDS AND OTHER BENZENEPOLYCARBOXYLIC ACIDS

This article discusses the benzenepolycarboxylic acids, their anhydrides, and their esters. Table 1 includes IUPAC nomenclature, common names, and CAS Registry Numbers for the benzenepolycarboxylic acids. These acids and anhydrides are highly stable. The carboxylic acid groups provide from two to six sites for reaction for a wide variety of products, mostly polymers and plasticizers.

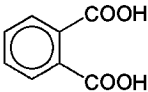
Table 1. Benzenepolycarboxylic Acids

IUPAC Name	Common name	CAS Registry Number	Formula weight
1,2-benzenedicarboxylic acid	phthalic acid	[88-99-3]	166.14
1,3-benzenedicarboxylic acid	isophthalic acid	[121-91-5]	166.14
1,4-benzenedicarboxylic acid	terephthalic acid	[100-21-0]	166.14
1,2,3-benzenetricarboxylic acid	hemimellitic acid	[569-51-7]	210.15
1,2,4-benzenetricarboxylic acid	trimellitic acid	[528-44-9]	210.15
1,3,5-benzenetricarboxylic acid	trimesic acid	[554-95-0]	210.15
1,2,3,4-benzenetetracarboxylic acid	mellophanic acid	[476-73-3]	254.16
1,2,3,5-benzenetetracarboxylic acid	prehnitic acid	[479-47-0]	254.16
1,2,4,5-benzenetetracarboxylic acid	pyromellitic acid	[89-05-4]	254.16
benzenepentacarboxylic acid		[1585-40-6]	298.17
benzenehexacarboxylic acid	mellitic acid	[517-60-2]	342.18

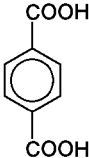
Phthalic anhydride (1) is the commercial form of phthalic acid (2). The worldwide production capacity for the anhydride was ca 3.5 × 10⁶ metric tons in 1993, and it was used in the manufacture of plasticizers (qv), unsaturated polyesters, and alkyd resins (qv) (see POLYESTERS, UNSATURATED). Sales of terephthalic acid (3) and its dimethyl ester are by far the largest of any of the benzenepolycarboxylic acids; 14.3 × 10⁶ t were produced in 1993. This is 80% of the total tonnage of all commercial forms of the benzenepolycarboxylic acids. Terephthalic acid is used almost exclusively for the manufacture of poly(ethylene terephthalate), which then is formed into textiles, films, containers, and molded articles. Isophthalic acid (4) and trimellitic anhydride (5) are commercial products, but their worldwide production capacities are an order of magnitude smaller than for terephthalic acid and its dimethyl ester. Isophthalic acid is used primarily in the production of unsaturated polyesters and as a comonomer in saturated polyesters. Trimellitic anhydride is used mainly to make esters for high performance poly(vinyl chloride) plasticizers. Trimesic acid (6), pyromellitic dianhydride (7), and hemimellitic acid (8) have specialized commercial applications. The rest of the benzenepolycarboxylic acids are not available commercially.



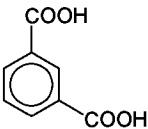
(1)



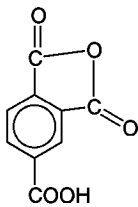
(2)



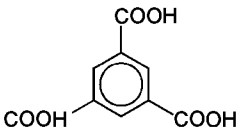
(3)



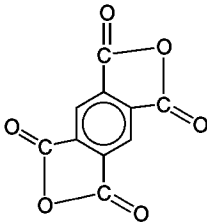
(4)



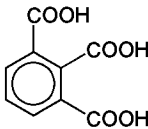
(5)



(6)



(7)



(8)

Physical Properties

The physical properties of the acids, the most important anhydrides, and the full methyl esters are summarized in Tables 2, 3, and 4. Detailed lists of physical properties for phthalic acid and its anhydride, terephthalic acid and dimethyl terephthalate, isophthalic acid, trimellitic acid and its anhydride, and pyromellitic acid and its dianhydride are provided under the sections describing these compounds.

Table 2. Physical Properties of Benzenepolycarboxylic Acids

Common name	Formula weight	Melting point, °C	Dissociation constants in aqueous solution, 25°C				ΔH_f° at 25°C kJ mol ^a	Solubility, g 100 g water		Reference ^b		
			p <i>K</i> ₁	p <i>K</i> ₂	p <i>K</i> ₃	p <i>K</i> ₄		at 25°C	at 100°C	(1)	(2)	(3)
phthalic	166.14	211 ^c	2.9	5.41			782	0.7	19.0	H 791	04526	5:2740
isophthalic	166.14	384	3.6	4.60			803	0.012	0.32	H 832	04527	3:1925
terephthalic	166.14	402 ^d	3.5	4.46			816	0.0017	0.033	H 841	04528	5:2949
hemimellitic ^e	210.15	197 ^c	2.8	4.20	4.87		1160	v sol	v sol	H 976	05571	1:331
trimellitic	210.15	238 ^c	2.5	3.84	5.20		1179	2.1	60	H 977	05572	1:331
trimesic	210.15	380	2.1	3.89	4.70		1190	0.24	6.4	H 978	05573	1:332
mellophanic	254.16	241 ^c	2.0	3.25	4.73	6.2	1562	sol	v sol	H 977	05529	1:331
prehnitic	254.16	238	2.3	3.51	4.44	5.8	1549			H 977	05530	1:331
pyromellitic	254.16	282	1.9	2.87	4.49	5.6	1571	1.5	>30	H 977	05531	1:331
benzenepentacarboxylic ^f	298.17	228	1.8	2.73	3.97	5.2	1930	sol	v sol	H 1006	05158	1:329
mellitic ^g	342.18	288 ^c	1.4	2.19	3.31	4.7	2299	sol	v sol	H 1008	04874	4:2067

^a To convert J to cal, divide by 4.184.
^b For Ref. 1, the letter and number refer to the section and paragraph; for Ref. 2, each compound is numbered; for Ref. 3, the numbers refer to the volume and page.
^c Decomposes at mp.

^d Sublimes.

^e Hemimellitic acid usually is handled as the dihydrate; formula wt = 246.18, mp = 191°C, decomposes.

^f $pK_s = 6.46$.

^g $pK_s = 5.89$; $pK_6 = 6.96$.

Table 3. Physical Properties of Anhydrides of the Benzenepolycarboxylic Acids

Common name	CAS Registry Number	Formula weight	Mp, °C	Bp, °C
phthalic anhydride	[85-44-9]	148.12	131	284.5
hemimellitic anhydride	[3786-39-8]	192.14	196	
trimellitic anhydride	[552-30-7]	192.14	168	390
mellophanic dianhydride	[4435-60-3]	218.13	198	
pyromellitic dianhydride	[89-32-7]	218.13	285	390
benzene-1,2,4,5-tetracarboxylic dianhydride-3-carboxylic acid	[59025-58-0]	262.14		
mellitic trianhydride	[4253-24-1]	288.14	320 dec	

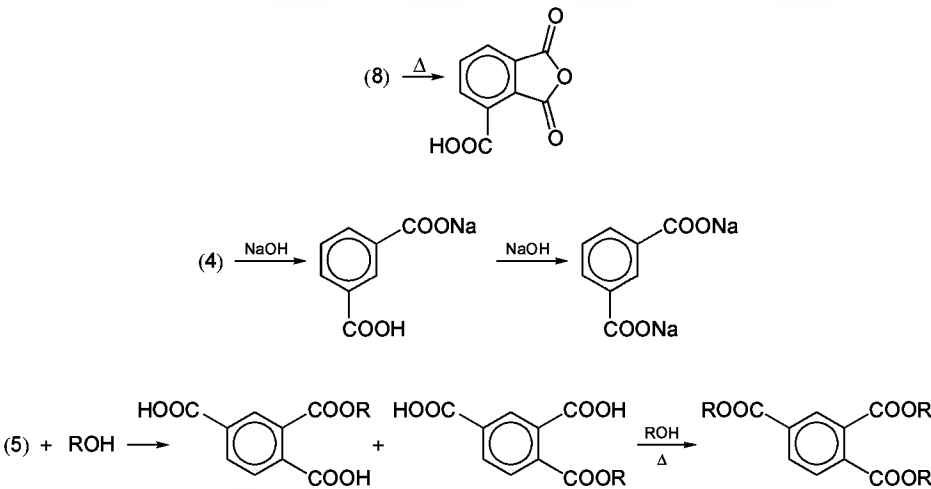
Table 4. Melting Points of the Full Methyl Esters of Benzenepolycarboxylic Acids

Common name	CAS Registry Number	Formula weight	Mp, °C
dimethyl phthalate	[131-11-3]	194.19	0–2
dimethyl isophthalate	[1459-93-4]	194.19	67
dimethyl terephthalate	[120-61-6]	194.19	141
trimethyl hemimellitate	[2672-57-3]	252.23	102
trimethyl trimellitate	[2459-10-1]	252.23	13
trimethyl trimesate	[2672-58-4]	252.23	146
tetramethyl mellophanate	[3451-02-3]	310.27	133
tetramethyl prehnitate	[3034-97-7]	310.27	112
tetramethyl pyromellitate	[635-10-9]	310.27	143 dec
pentamethyl benzenepentacarboxylate	[3327-06-8]	368.30	148
hexamethyl mellitate	[6237-59-8]	426.34	188

Chemical Properties

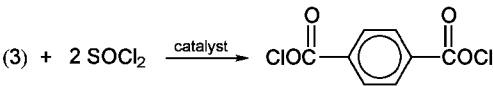
The chemistry of benzenecarboxylic acids generally is the same as that of other carboxylic acids, which can be converted into esters, salts, acid chlorides, and anhydrides. Each carboxyl group can react separately, so that compounds in which carboxyl groups are converted into different derivatives can be prepared. Because there are aromatic hydrogens available in most of these acids, they also undergo reactions characteristic of the benzene nucleus. Some of the anhydrides have characteristic reactions.

Reactions of the Carboxyl Groups. Carboxyl groups in the ortho position spontaneously form a strainless five-membered ring when heated to give anhydrides as shown for (8). Salts and esters (4) are readily formed as shown for (4) and (5), respectively.

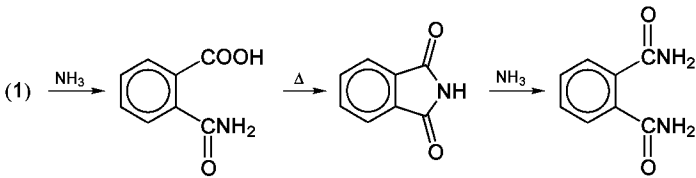


Because each carboxyl group reacts at a different rate with alcohol, a mixed diester or triester can be prepared if the reactions of the acid with ROH and R'OH are carried out sequentially.

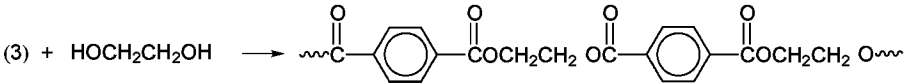
Acid chlorides can be formed by reaction with thionyl chloride:



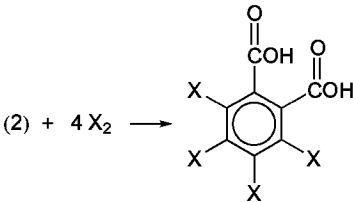
The carboxylic acids react with ammonia and with primary amines in a similar manner:



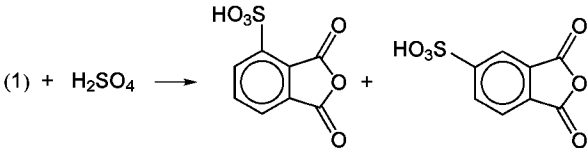
Condensation polymerization also occurs and essentially is the sole reason for the commercial existence of terephthalic acid:



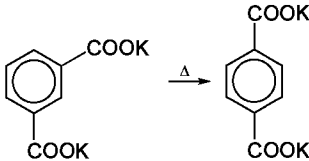
Reactions of the Benzene Ring. The benzene rings of the benzenepolycarboxylic acids undergo halogenation:



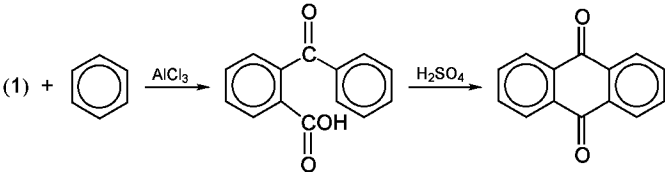
and sulfonation:



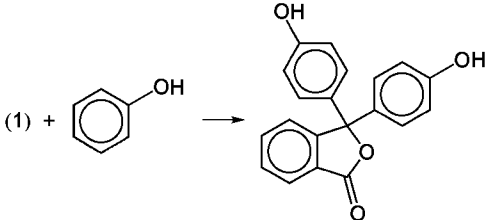
When mixed phthalic acids are converted to their dipotassium salts, they can be thermally or catalytically rearranged to the para isomer. This rearrangement is known as the Henkel reaction:



Other Reactions. Phthalic anhydride (1) undergoes condensation to form anthraquinone derivatives:



Also, (1) undergoes condensation with phenol to form phenolphthalein:



Phthalic anhydride reacts with urea and metal diacetates to form metal phthalocyanines (qv).

Phthalic Acid and Phthalic Anhydride

The first of the benzene polycarboxylic acids to become a commercial product was phthalic acid, mostly in the form of the anhydride. The anhydride is obtained by the catalytic vapor-phase air oxidation of *o*-xylene or naphthalene. The IUPAC name of phthalic anhydride is 1,3-isobenzofurandione [85-44-9].

Physical and Chemical Properties. Tables 5, 6, and 7 list some of the physical and chemical properties of phthalic acid and its anhydride.

Table 5. Physical Constants of Phthalic Acid and Phthalic Anhydride

Property	Phthalic acid	Phthalic anhydride
mp, °C	211 dec	131
bp, °C		284.5 sub
triple point, °C		131
heat of vaporization at 131°C, kJ mol ³		65.3
specific gravity at 4°C	1.593	1.527

specific heat, J (kgK) ^a			
at 300°C	1138		1009
200°C	800		763
90°C	433		422
heat of combustion at 25°C, kJ mol ^a	3224	3259	
heat of formation at 25°C, kJ mol ^a	784	460	
heat of sublimation at 131°C, kJ mol ^a			88.70
heat of fusion at 131°C, kJ mol ^a			22.93

^a To convert J to cal, divide by 4.184.

Table 6. Physical Properties of Liquid Phthalic Anhydride

Temperature, °C	Vapor presure, kPa ^a	Density, g cm ³	Viscosity, mPas(cP)
132°C	0.8		1.19
135°C	0.9	1.215	
140°C	1.2	1.208	
160°C	2.7	1.188	
180°C	5.5	1.166	
197°C	10	1.278	0.94
200°C	10.7	1.145	
220°C		1.131	0.55
240°C		1.103	
284.5°C	101		

^a To convert kPa to mm Hg, multiply by 7.5.

Table 7. Solubilities of Phthalic Acid and Phthalic Anhydride^a

Solvent	Phthalic acid	Phthalic anhydride
water		
at 25°C	0.7	0.6
at 100°C	19.0	16.5
at 150°C	200	147
ethanol at 18°C	11.7	
diethyl ether at 15°C	0.7	
carbon disulfide at 20°C		0.7
formic acid (95 wt %) at 20°C		4.7
pyridine at 20–25°C		80
acetic acid (glacial) at 100°C	12.0	

^a g/100 g solvent.

Manufacture and Processing. Until World War II, phthalic acid and, later, phthalic anhydride, were manufactured primarily by liquid-phase oxidation of suitable feedstocks. The favored method was BASF's oxidation of naphthalene [91-20-3] by sulfuric acid in the presence of mercury salts to form the anhydride. This process was patented in 1896. During World War I, a process to make phthalic anhydride by the oxidation of naphthalene in the vapor phase over a vanadium and molybdenum oxide catalyst was developed in the United States (5). Essentially the same process was developed independently in Germany, with U.S. patents being granted in 1930 and 1934 (6,7).

Naphthalene (qv) from coal tar continued to be the feedstock of choice in both the United States and Germany until the late 1950s, when a shortage of naphthalene coupled with the availability of xylenes from a burgeoning petrochemical industry forced many companies to use *o*-xylene [95-47-6] (8). Air oxidation of 90% pure *o*-xylene to phthalic anhydride was commercialized in 1946 (9,10). An advantage of *o*-xylene is the theoretical yield to phthalic anhydride of 1.395 kg/kg. With naphthalene, two of the ten carbon atoms are lost to carbon oxide formation and at most a 1.157-kg/kg yield is possible. Although both are suitable feedstocks, *o*-xylene is overwhelmingly favored. Coal-tar naphthalene is used in some cases, eg, where it is readily available from coke operations in steel mills (see STEEL). Naphthalene can be produced by hydrodealkylation of substituted naphthalenes from refinery operations (8), but no refinery-produced naphthalene is used as feedstock. Alkyl naphthalenes can be converted directly to phthalic anhydride, but at low yields (11,12).

Fixed-Bed Vapor-Phase Oxidation of *o*-Xylene. Well in excess of 90% of the phthalic anhydride produced is obtained by oxidizing *o*-xylene in the vapor phase over a fixed bed of catalyst. In the 1960s, there were two types of fixed-bed processes, low temperature/low space velocity, and high temperature/high space velocity. Catalyst development resulted in higher allowable space velocities for the low temperature case while high yields were maintained. Consequently, use of the low temperature process which runs at <400°C has predominated. Processes of this type are operated by companies listed in Table 9 as well as BASF, Alusuisse, Huels, BP Chemical, and Bayer. A typical flow sheet is shown in Figure 1. A commercially viable plant must operate at high selectivity of at least 75 mol % with a feed *o*-xylene concentration of 60 g/m³ (13). This concentration is above the lower explosion limit of 43 g/m³ (14). The catalyst should last at least three years (13).

Table 8. Estimated World Capacities for Phthalic Anhydride, 10³ t

Location	Year		
	1976	1988	1993
North America	440	560	455
Western Europe	610	860	950
Far East	280	760	1090
Eastern Europe/Mid-East	200	750	660

South America	35	305	350
Total	1565	3235	3505

Table 9. U.S. Capacities for Phthalic Anhydride, 10³ t

Company	Year	
	1980	1993
Exxon	59	118
Aristech	91	104
Stepan	91	79
Koppers	147	79
Sterling	0	75
BASF	68	0
Monsanto	118	0
Chevron	23	0
Allied	18	0
Total	615	455

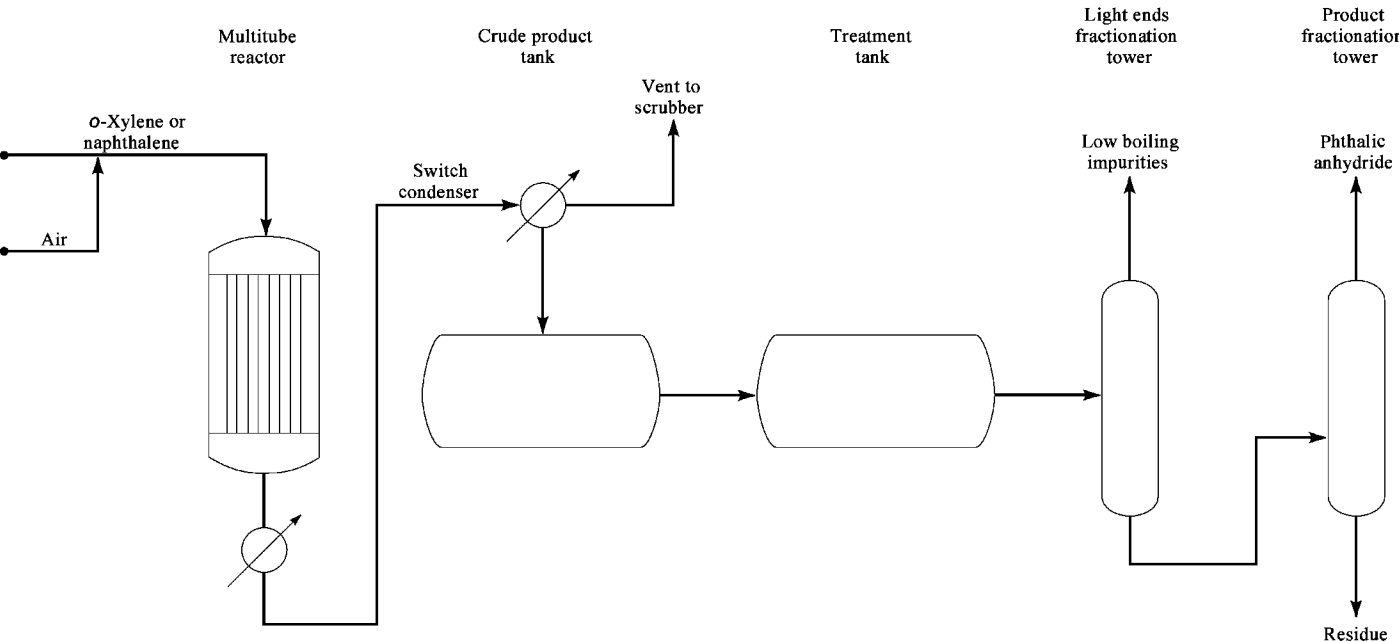


Fig. 1. A typical fixed-bed process for phthalic anhydride.

Some reactors are designed specifically to withstand an explosion (14). The multitube fixed-bed reactors typically have ca 2.5-cm inside-diameter tubes, and heat from the highly exothermic oxidation reaction is removed by a circulating molten salt. This salt is a eutectic mixture of sodium and potassium nitrate and nitrite. Care must be taken in reactor design and operation because fires can result if the salt comes in contact with organic materials at the reactor operating temperature (15). Reactors containing over 20,000 tubes with a 45,000-ton annual production capacity have been constructed.

The catalyst combines two essential ingredients found in earlier catalysts, vanadium oxide and titanium dioxide, which are coated on an inert, nonporous carrier in a layer 0.02- to 2.0-mm thick (13,16). Other elements such as phosphorus are also used. Ring-shaped supports are used instead of spherical supports to give longer catalyst life, less pressure drop though the reactor, and higher yields (17,18). Half rings are even better and allow more catalyst to be loaded (18).

There are thermal gradients along the catalyst tube, and a temperature maximum or hot spot develops which tends to place limits in the feed rate owing to possible catalyst damage and to decrease the yield. Various methods are used to reduce this hot spot, including two or more catalysts in series to limit the reaction initially. This use of catalysts could be higher rubidium or potassium concentrations in the first bed, with higher phosphorus levels in the second (19). Different phosphorus levels may also be used (20). Placing varying concentrations of inert packing along the tube length is also possible (18,21,22). Another method is to have two temperature zones, the first zone being cooler than the second (23).

Fixed-Bed Vapor-Phase Oxidation of Naphthalene. A silica gel or silicon carbide support is used for catalyst involved in the oxidation of naphthalene. The typical naphthalene oxidation catalyst is a mixture of vanadium oxide and alkali metal sulfate on the silica support. Some changes, such as the introduction of feed vaporizers, are needed to handle a naphthalene feed (14), but otherwise the equipment is the same.

Fluidized-Bed Vapor-Phase Oxidation. Fluidized-bed reactors were developed for oxidation of naphthalene to phthalic anhydride and started operation in 1945 (24). In the 1990s, there are very few naphthalene-feed fluidized-bed plants that are operational, located in the Far East only, and even there possibly only one. There are none using *o*-xylene. The silica gel catalyst support used in the fluidized bed interferes with the oxidation of *o*-xylene (25). Conversely, the titanium dioxide support used in *o*-xylene oxidation does not stand up well in the reactor. As with all fluidized-bed processes, catalyst attrition must be considered and minimized. Commercial plants use a low activity catalyst on silica gel at 340–385°C operating temperature. In spite of the seemingly low activity and temperature, all the naphthalene is converted in the bed, although not necessarily to phthalic anhydride (25). The intense mixing in fluidized beds offers several advantages, as illustrated by the complete oxidation noted above. It also minimizes problems of flammability, especially with coal-tar naphthalenes, which may contain pyrophoric substances (26). The agitated catalyst particles quickly dissipate heat, and some catalyst fines in the reactor effluent further prevent buildup of potentially dangerous substances. The capacity of one fluidized-bed process can be adjusted by the operating pressure, which, in turn, changes the time in which the feedstock is in contact with the catalyst (27).

Liquid-Phase Oxidation of *o*-Xylene. A commercial plant which oxidized *o*-xylene in the liquid phase using acetic acid solvent and a cobalt–manganese–bromine catalyst system was operated in France starting about 1965 but was shut down in the late 1970s. This type of process is described in the section on terephthalic acid. The reaction produces phthalic acid, which is then dehydrated to the anhydride. Yields from this process are high; the yield to phthalic anhydride is at least 90 mol % vs 80 mol % or less for vapor-phase oxidation. Capital costs are high owing to the metallurgy required.

Phthalic Anhydride Recovery and Purification. The accepted method of recovering phthalic anhydride from vapor-phase oxidation

processes is first to recover some of the heat of oxidation, eg, by using the reactor effluent to generate steam. Then the vapors are passed through automatically cycled switch condensers where up to 99.5% of the phthalic anhydride solidifies on cooled, finned tubes. Maintaining the condenser temperature above the 131°C phthalic anhydride melting point results in unacceptable product losses. Choice of temperature is critical, because only about 1 vol % of the reactor effluent is phthalic anhydride, 8 vol % water (13). Water condensation must be avoided to prevent formation of the acid from the anhydride. At a preset time, the switch condenser in use is taken off line and another one started. The off-line condenser is then charged with hot oil to melt the anhydride into a storage tank (28). In fluidized-bed processes, the phthalic anhydride concentration is high enough to first condense some 50% of it as a liquid before going to the switch condensers (24). Vent gases from the switch condensers contain maleic anhydride, citraconic acid, and benzoic acid as well as phthalic anhydride. The gas can be scrubbed with water or incinerated, either thermally or catalytically (23), before being released into the atmosphere. The water used to scrub the gases contains these mixed acids, and maleic anhydride can be recovered (29). However, the economics prohibit recovery in view of the high cost of capital equipment vs the small amount of maleic anhydride obtained.

The crude phthalic anhydride is subjected to a thermal pretreatment or heat soak at atmospheric pressure to complete dehydration of traces of phthalic acid and to convert color bodies to higher boiling compounds that can be removed by distillation. The addition of chemicals during the heat soak promotes condensation reactions and shortens the time required for them. Use of potassium hydroxide and sodium nitrate, carbonate, bicarbonate, sulfate, or borate has been patented (30). Purification is by continuous vacuum distillation, as shown by two columns in Figure 1. The most troublesome impurity is phthalide (1(3)-isobenzofuranone), which is structurally similar to phthalic anhydride. Reactor and recovery conditions must be carefully chosen to minimize phthalide contamination (31). Phthalide [87-41-2] is also reduced by adding potassium hydroxide during the heat soak (30).

Production, Storage, and Shipment. All U.S. producers have at least a 75,000-t annual capacity. There are some small plants and small producers in Europe with less than 20,000-t capacity. These are gradually being shut down, and any new plant would have to be large in order to reach economies of scale in the present market (ca 1995), which is highly competitive.

Storage and shipment are preferably molten. Insulated and heated tanks are used, as are insulated rail tank cars and tank trucks. In the United States, rail tank cars can hold about 84 t, and trucks about 20 t. The molten form can be handled and pumped in bulk form, and as a result is priced lower than the solid. Solid phthalic anhydride is available as flake in 1-t and 0.5-t super sacks, and 22.7-kg multiwall bags.

Economic Aspects. The estimated world production of phthalic anhydride was 2.7×10^6 t in 1993 (32), well below the 3.5×10^6 -t capacity. As seen in Table 8, production has shifted strongly from the developed to the developing areas of the world since the mid-1970s, especially to the Far East. Western Europe includes everything except the former Soviet Union. Eastern Europe Mid-East then includes the former Soviet Union as well as Pakistan and India, a region which has also experienced high growth. Closing of many old, small, and environmentally less clean plants is expected to continue in the late 1990s.

U.S. capacity is 455×10^3 t among five producers, as shown in Table 9 (33). All U.S. producers use *o*-xylene as a feedstock, although Koppers can switch to coal-tar naphthalene. Europe has 20 producers operating 25 plants, 18 of which use *o*-xylene. Demand growth in North America has averaged less than 3% annually since 1981. Prices in 1992 were \$0.77–0.95 kg for molten, and \$0.82–1.01 kg for flake (33).

Specifications and Standards. Typical specifications for phthalic anhydride are given in Table 10. All specifications are measures of purity. Solidification point is a sensitive indicator of absolute purity, and is a key specification. Another key specification is molten color stability, which is the color after being held at 250°C for two hours. This test ensures acceptable color after shipment in molten form and detects the presence of impurities that can cause discoloration at elevated temperatures. Phthalic acid level is a monitor of how well moisture has been excluded during storage and shipment.

Table 10. Specifications for Phthalic Anhydride

Property	Specification ^a	Test method
freezing point, °C	130.9 ^b	cryoscopy
molten color		
initial, at 250°C, APHA	20	colorimetry
stability, 2 h at 250°C, APHA	60	colorimetry
phthalic anhydride, %	99.7 ^b	gas chromatography
phthalic acid, %	0.2	titration
maleic acid, %	0.05	gas chromatography
naphthoquinone, %	0.0002	gas chromatography

^a Values are maximum unless otherwise noted.

^b Minimum.

Analytical and Test Methods. Measurement of the solidification point using a highly sensitive thermometer and of APHA color by comparison of molten samples to APHA standards is straightforward. Specific impurities are measured by gas chromatography. A nonaqueous titration is used to determine phthalic acid content.

Health and Safety Factors. Phthalic anhydride is a severe irritant to the eyes, respiratory tract, and skin, especially to moist tissue (34–36). The solid may burn skin tissue if it is in contact with it for a significant amount of time. Repeated exposure may result in asthma, irritation of mucous membranes, and diseases of the respiratory tract and digestive organs (37,38). Contact with skin or the eyes should be followed immediately by washing with large quantities of water. Goggles, a face shield, and heavy leather gauntlets should be worn by workers handling phthalic anhydride. Phthalic anhydride and phthalic acid are toxic, though not strongly so; the LD₅₀ by oral ingestion for rats is 4 g kg for the anhydride and 7.9 g kg for the acid (39).

There are explosion hazards with phthalic anhydride, both as a dust or vapor in air and as a reactant. Table 11 presents explosion hazards resulting from phthalic anhydride dust or vapor (40,41). Preventative safeguards in handling solid phthalic anhydride have been reported (15). Water, carbon dioxide, dry chemical, or foam may be used to extinguish the burning anhydride. Mixtures of phthalic anhydride with copper oxide, sodium nitrite, or nitric acid plus sulfuric acid above 80°C explode or react violently (39).

Table 11. Flammability and Explosivity of Phthalic Anhydride

Property	Value
	<i>Dust cloud</i>
explosibility index	>10
ignition sensitivity	13.8
ignition temperature, °C	650
minimum igniting energy, J ^a	0.015
minimum explosive concentration, g L	0.015
relative flammability, % inert	
by furnace ignition	>90
by spark ignition	90
limiting oxygen concentration, %	
by furnace ignition	11
by spark ignition	14

explosive severity	1.6
dust concentration, g L	
0.1 g L dust concentration	269
0.5 g L dust concentration	496
maximum pressure, kPa ^b	
rate of pressure rise, 10 ³ kPa s ^b	
6.9, average	9.7
9.0, max	29.0
<i>Molten liquid</i>	
flash point, °C	
open cup	165
closed cup	151
autoignition temperature, °C	584
<i>Vapor</i>	
explosive limits in air at 140–285°C, vol %	
lower	1.7
upper	10.5

^a To convert J to cal, divide by 4.184.

^b To convert kPa to psi, multiply by 0.145.

Uses. Phthalic anhydride is used mainly in plasticizers, unsaturated polyesters, and alkyd resins (qv). Phthalic plasticizers consume 54% of the phthalic anhydride in the United States (33). The plasticizers (qv) are used mainly with poly(vinyl chloride) to produce flexible sheet such as wallpaper and upholstery fabric from normally rigid polymers. The plasticizers are of two types: diesters of the same monohydric alcohol such as dibutyl phthalate, or mixed esters of two monohydric alcohols. The largest-volume plasticizer is di(2-ethylhexyl) phthalate [117-81-7] which is known commercially as dioctyl phthalate (DOP) and is the base to which other plasticizers are compared. The important phthalic acid esters and their physical properties are listed in Table 12. The demand for phthalic acid in plasticizers is naturally tied to the growth of the flexible poly(vinyl chloride) market which is large and has been growing steadily.

Table 12. Properties of Phthalic Acid Esters

Ester	CAS Registry Number	Mp, °C	Bp, °C	Specific gravity °C
dimethyl phthalate	[131-11-3]	0–2	282	1.192 20°C
diethyl phthalate	[84-66-2]	4	296	1.118 20°C
diallyl phthalate	[131-17-9]	65	156–175 ^a	1.120 20°C
dibutyl phthalate	[84-74-2]	40	340	1.042 25°C
butyl cyclohexyl phthalate	[84-64-0]		189–222 ^b	1.076 25°C
diamyl phthalate	[131-18-0]	< 55	342	1.022 25°C
butyl benzyl phthalate	[85-68-7]		370	1.111 25°C
dicyclohexyl phthalate	[84-61-1]	58–65	212–218 ^b	1.148 20°C
butyl octyl phthalate	[84-78-6]	< 50	225 ^b	0.993 25°C
butyl decyl phthalate	[89-19-0]	50	220 ^b	0.991 25°C
di(<i>n</i> -octyl) phthalate	[117-84-0]	25	220–248 ^a	0.978 20°C
diisooctyl phthalate	[27554-26-3]	< 50	228–239 ^b	0.986 20°C
di(2-ethylhexyl) phthalate	[117-81-7]	46	231 ^b	0.986 20°C
<i>n</i> -octyl <i>n</i> -decyl phthalate	[119-07-3]	28	250 ^b	0.970 25°C
isooctyl isodecyl phthalate	[42343-35-1]	48	235–248 ^a	0.967 25°C
diisodecyl phthalate	[19269-67-1]	48	255 ^b	0.961 25°C

^a At 533 Pa (4.0 mm Hg).

^b At 666 Pa (5.0 mm Hg).

The second largest use at 21% is for unsaturated polyester resins, which are the products of polycondensation reactions between molar equivalents of certain dicarboxylic acids or their anhydrides and glycols. One component, usually the diacid or anhydride, must be unsaturated. A vinyl monomer, usually styrene, is a diluent which later serves to fully cross-link the unsaturated portion of the polycondensate when a catalyst, usually a peroxide, is added. The diacids or anhydrides are usually phthalic anhydride, isophthalic acid, and maleic anhydride. Maleic anhydride provides the unsaturated bonds. The exact composition is adjusted to obtain the required performance. Resins based on phthalic anhydride are used in boat hulls, tubs and spas, construction, and synthetic marble surfaces. In most cases, the resins contain mineral or glass fibers that provide the required structural strength. The market for the resins tends to be cyclical because products made from them sell far better in good economic times (see POLYESTERS, UNSATURATED).

The manufacture of alkyd resins (qv), which are obtained by the reactions of polybasic acids or anhydrides, polyhydric alcohols, and fatty oils and acids, consumes about 17% of the phthalic anhydride demand. While materials such as maleic anhydride, isophthalic acid, and fumaric acid can also be used, phthalic anhydride is the most important. The resin provides a binder for coatings that are applied for either protection or decoration. Air quality concerns have put alkyd resins under pressure from water-based coatings which do not emit organic vapors upon drying.

Derivatives. The remaining uses are better considered as derivatives of phthalic anhydride and consume less than 10% of the demand, but they provide a diverse group of products. Anthraquinone (qv), a starting reagent for a number of dye intermediates, is manufactured by heating phthalic acid and benzene with sulfuric acid as a catalyst. Use of chlorobenzene instead of benzene results in 2-chloroanthraquinone, a dye. Examples of dyes derived from phthalic anhydride are phthalocyanine blues, quinoline yellow, and anthracene brown (see DYES AND DYE INTERMEDIATES). Another use for phthalic anhydride is in the production of isatoic anhydride [118-48-9], a raw material used in the production of saccharin. Tetrachloro- and tetrabromophthalic anhydrides are manufactured by the reaction of phthalic anhydride with chlorine and bromine, respectively, at high temperatures. The halogenated forms impart fire resistance to polyester resins, polyurethane foams, and surface coatings. Phenolphthalein [77-09-8] is the condensation product of phthalic anhydride and phenol in the presence of a dehydrating agent, and is a pH indicator and laxative. Pesticides and anthranilic acid [118-92-3] can be made from phthalimide which is in turn produced from phthalic anhydride.

Terephthalic Acid and Dimethyl Terephthalate

Purified terephthalic acid and dimethyl terephthalate are used as raw materials for the production of saturated polyesters. During 1993, the combined worldwide production of purified terephthalic acid plus dimethyl terephthalate exceeded 14×10^6 t (42), which is 80% of the total benzenepolycarboxylic acid production. Terephthalic acid is also produced in technical or crude grades which are not pure enough for manufacture of poly(ethylene terephthalate). In almost all cases, the technical-grade material is immediately converted to purified terephthalic acid or dimethyl terephthalate, which together are the articles of commerce.

Physical and Chemical Properties. Tables 13, 14, 15, 16, and 17 contain the more important physical and some chemical properties of terephthalic acid and dimethyl terephthalate.

Table 13. Physical Constants of Terephthalic Acid and Dimethyl Terephthalate

Property	Terephthalic acid		Dimethyl terephthalate
mp, °C	427 ^a		140.65
bp, °C			284
triple point, °C	427		140.65
heat of fusion, kJ mol ^b			31.6
heat of vaporization, kJ mol ^b			57.3
cryoscopic constant, mol ⁻¹ °C			2.28
specific gravity at 25°C	1.522		1.283
specific heat, J (kgK) ^b	1202	1400 ^c	
sublimation point, °C	404		
heat of combustion at 25°C, kJ mol ^b	3198	4685	
heat of formation at 25°C, kJ mol ^b	816	711	
heat of sublimation, kJ mol ^b	142		88.4

^a Sealed tube.
^b To convert J to cal, divide by 4.184.
^c At 50°C.

Table 14. Vapor Pressures of Terephthalic Acid^a and Dimethyl Terephthalate^b

Temperature, °C		
Terephthalic acid	Dimethyl terephthalate	Pressure, kPa ^c
303°C	148°C	1.3
353°C	210°C	13.3
370°C	233°C	26.7
387°C	258°C	53.3
404°C	284°C	101.3

^a Solid.
^b Liquid.
^c To convert kPa to mm Hg, multiply by 7.5.

Table 15. Physical Properties of Liquid and Vapor Dimethyl Terephthalate

Property	Temperature, °C		
	150°C	200°C	250°C
	<i>Liquid</i>		
density, g cm ³	1.068	1.022	0.976
viscosity, mPas(−c P)	0.965	0.595	0.405
specific heat, J (kgK) ^a	1921	2030	
	<i>Vapor</i>		
viscosity, mPas(−cP)	0.0086	0.0096	0.0107
specific heat, J (kgK) ^a	1290	1390	1500

^a To convert J (kgK) to cal (g°C), divide by 4184.

Table 16. Solubilities of Terephthalic Acid^a

Solvent	Temperature, °C				
	25°C	120°C	160°C	200°C	240
in water	0.0017	0.07	0.37	1.8	8.5
in glacial acetic acid	0.013	0.17	0.50	1.5	4.3
in methanol	0.1	2.1	2.9	15.0	
in dimethylformamide	6.7				
in dimethyl sulfoxide	19.0				

^a g 100 g solvent.

Table 17. Solubilities of Dimethyl Terephthalate^a

Solvent	At 25°C	At 60°C	At bp of solvent, (bp, °C)
in ethylene glycol	0.8		
in diethylene glycol			400 245°C
in methanol	1.0	5.7	
in carbon tetrachloride	1.5	3.6	25 77°C
in ethyl ether	1.6		
in acetone			25 56°C
in methyl ethyl ketone	1.6	12.5	
in benzene	2.0	14.0	
in ethyl acetate	3.5	16.0	
in butyl acetate			400 120°C
in toluene	4.3	10.4	>100 111°C
in ethylenediamine			25 117°C
in dioxane	7.5	28.5	
in chloroform	10.0	23.0	

^a g 100 g solvent.

Manufacture and Processing. Terephthalic acid and dimethyl terephthalate did not become large-volume industrial chemicals until after World War II. Imperial Chemical Industries in the United Kingdom in 1949 and Du Pont in the United States in 1953 commercialized fibers made from poly(ethylene terephthalate). Dimethyl terephthalate and ethylene glycol were the comonomers used by both companies (see FIBERS, POLYESTER).

Initial production of the dimethyl terephthalate started with the oxidation of *p*-xylene to terephthalic acid using nitric acid; both companies reportedly used similar technology (43–45). Versions of the nitric acid oxidation process, which has been abandoned commercially, involved the use of air in the initial oxidation step to reduce the consumption of nitric acid (44,46,47). The terephthalic acid was then esterified with methanol to produce dimethyl terephthalate, which could be purified by distillation to the necessary degree (48).

p-Xylene [106-42-3] is still the only feedstock used for either product. However, purified terephthalic acid has replaced dimethyl terephthalate as the leading terephthalate source for poly(ethylene terephthalate). About 75% of poly(ethylene terephthalate) production worldwide is based on the purified acid, and this percentage is increasing as virtually no new dimethyl terephthalate capacity is being built. Terephthalic acid offers distinct cost and product quality advantages to the manufacturer of poly(ethylene terephthalate). This has led to its use in almost all new polymer plants. Specifically, polyester processes that are based on dimethyl terephthalate must have equipment for recovery of methanol, which is the by-product of the transesterification with ethylene glycol. Use of the pure acid produces water as a by-product of direct esterification. In addition, terephthalic acid provides a higher yield of polyester per kilogram of starting feedstock. While excess ethylene glycol is needed in both processes, the percentage of excess needed is much smaller using terephthalic acid, so less in the way of recovery and recycle process equipment is required. The need for transesterification catalysts with dimethyl terephthalate introduces metals into the polyester which can cause undesirable side reactions (49,50).

Purified terephthalic acid became commercially available from Amoco Chemical Co. in 1965, by which time a considerable polyester industry based on dimethyl terephthalate had already developed. The Amoco process involves purification of crude terephthalic acid by a separate step to attain the high product purity required for polyester manufacture. The Amoco technology is the most-used worldwide, but other processes have been developed and are operating commercially.

Technical-Grade Terephthalic Acid. All technical-grade terephthalic acid is produced by catalytic, liquid-phase air oxidation of *p*-xylene. Several processes have been developed, but they all use acetic acid as a solvent and a multivalent heavy metal or metals as catalysts. Cobalt is always used. In the most popular process, cobalt and manganese are the multivalent heavy-metal catalysts and bromine is the renewable source for free radicals (51,52). This catalyst system is used in about 70% of the *p*-xylene oxidations, and the percentage is increasing as new plants almost invariably employ it. Process conditions are highly corrosive owing to the acetic acid and bromine, and titanium must be used in contact with some parts of the process.

Figure 2 is a typical flow sheet for the process using technology originally developed and licensed by Amoco and Mitsui Petrochemical. Acetic acid, air, *p*-xylene, and catalyst are fed continuously into an oxidation reactor that is maintained at 175–225°C and 1500–3000 kPa (~15–30 atm). Air is added in amounts in excess of stoichiometric requirements to minimize formation of by-products. The oxidation is exothermic to the extent of 2×10^8 J/kg of *p*-xylene reacted, and this heat is removed by allowing the acetic acid solvent to boil. The vapor is condensed and refluxed to the reactor, and this sets the temperature–pressure relationship. The condensing vapor is used to generate steam, which is employed as a heat source in other parts of the process. Two moles of water are formed per mole of *p*-xylene reacted. The residence time is 30 min–2 h depending on the process. More than 98% of the *p*-xylene is converted and the yield to terephthalic acid is at least 95 mol % in modern plants. Further, this is on a once-through basis. The near exclusive selection of this technology for all new capacity follows from these yield and conversion values.

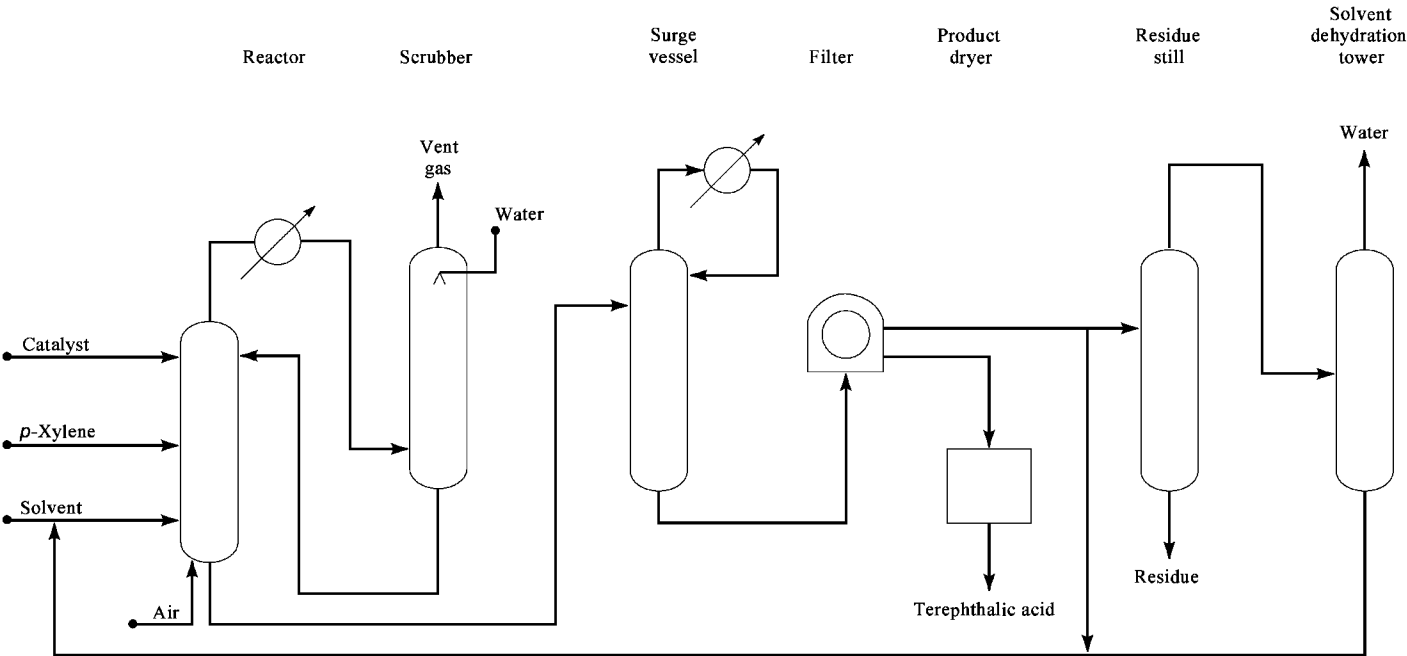


Fig. 2. Terephthalic acid production by catalytic, liquid-phase air oxidation of *p*-xylene.

The effluent from the reactor is a slurry of terephthalic acid because it dissolves to a limited extent in almost all solvents, including the acetic acid–water solvent used here. This slurry passes through a surge vessel that operates at a lower pressure than the reactor. More terephthalic acid crystallizes and the slurry is then ready to be processed at close to atmospheric conditions. The terephthalic acid crystals are recovered by filtration, washed, dried, and conveyed to storage, from which they are in turn fed to the purification step.

This is called a technical or crude grade of terephthalic acid, but the purity is typically greater than 99%. It is not, however, pure enough for the poly(ethylene terephthalate) made from it to reach the required degree of polymerization. The main impurity is 4-formylbenzoic acid [619-66-9], which is incompletely oxidized *p*-xylene and is monofunctional with regard to esterification. 4-Formylbenzoic acid is usually referred to as 4-carboxybenzaldehyde (4-CBA) in the industry.

Water formed in the reaction as well as some undesirable by-products must be removed from the acetic acid solvent. Therefore, mother liquor from the filter is purified in a residue still to remove heavies, and in a dehydration tower to remove water. The purified acetic acid from the bottom of the dehydration tower is recycled to the reactor. The water overhead is sent to waste treatment, and the residue still bottoms can be processed for catalyst recovery. Alternatively, some mother liquor from the filter can be recycled directly to the reactor.

Cleanup of waste streams from the process has been highly developed and is widely practiced commercially. The nitrogen and unused oxygen from the oxidation reactor are scrubbed to recover and reuse valuable components. The gas can then pass to a catalytic oxidation step, followed by a second scrubber to remove trace components and thereby meet the most demanding environmental requirements (53). Wastewater is treated by aerobic oxidation with specially acclimated bacteria, again in order to meet any environmental requirements. Alternatively, an anaerobic wastewater treatment process has been developed and commercially installed which produces much less waste sludge, requires less energy, and in addition produces methane, which can be burned for energy recovery (54).

Acetaldehyde can be used as an oxidation-promoter in place of bromine. The absence of bromine means that titanium metallurgy is not required. Eastman Chemical Co. has used such a process, with cobalt as the only catalyst metal. In that process, acetaldehyde is converted to acetic acid at the rate of 0.55–1.1 kg/kg of terephthalic acid produced. The acetic acid is recycled as the solvent and can be isolated as a by-product. Reaction temperatures can be low, 120–140°C, and residence times tend to be high, with values of two hours or more (55). Recovery of dry terephthalic acid follows steps similar to those in the Amoco process. Eastman has abandoned this process in favor of a bromine promoter (56). Another oxidation promoter which has been used is paraldehyde (57), employed by Toray Industries. This leads to the coproduction of acetic acid. 2-Butanone has been used by Mobil Chemical Co. (58).

Henkel Rearrangement of Benzoic Acid and Phthalic Anhydride. Henkel technology is based on the conversion of benzenecarboxylic acids to their potassium salts. The salts are rearranged in the presence of carbon dioxide and a catalyst such as cadmium or zinc oxide to form dipotassium terephthalate, which is converted to terephthalic acid (59–61). Henkel technology is obsolete and is no longer practiced, but it was once commercialized by Teijin Hercules Chemical Co. and Kawasaki Kasei Chemicals Ltd. Both processes followed a route starting with oxidation of naphthalene to phthalic anhydride. In the Teijin process, the phthalic anhydride was converted sequentially to monopotassium and then dipotassium *o*-phthalate by aqueous recycle of monopotassium and dipotassium terephthalate (62). The dipotassium *o*-phthalate was recovered and isomerized in carbon dioxide at a pressure of 1000–5000 kPa (~10–50 atm) and at 350–450°C. The product dipotassium terephthalate was dissolved in water and recycled as noted above. Production of monopotassium *o*-phthalate released terephthalic acid, which was filtered, dried, and stored (63,64).

Mitsubishi Chemical Industries, Ltd. practiced a Henkel II technology starting with toluene to produce benzoic acid. Reaction of benzoic acid with potassium hydroxide resulted in potassium benzoate, which was subjected to a disproportionation reaction to produce dipotassium terephthalate and benzene. Dipotassium terephthalate reacted with sulfuric acid, and the resulting terephthalic acid was recovered by filtration and drying (65,66). Here, dipotassium sulfate was the by-product.

Polymer-Grade Dimethyl Terephthalate.

Hercules/Dynamit Nobel Process. On a worldwide basis, the Hercules Inc. Dynamit Nobel AG process is the dominant technology for the production of dimethyl terephthalate; the chemistry was patented in the 1950s (67–69). Modifications in commercial practice have occurred over the years, with several variations being practiced commercially (70–72). The reaction to dimethyl terephthalate involves four steps, which alternate between liquid-phase oxidation and liquid-phase esterification. Two reactors are used. First, *p*-xylene is oxidized with air to *p*-toluic acid in the oxidation reactor, and the contents are then sent to the second reactor for esterification with methanol to methyl *p*-toluate. The toluate is isolated by distillation and returned to the first reactor where it is further oxidized to monomethyl terephthalate, which is then esterified in the second reactor to dimethyl terephthalate.

Figure 3 is a flow diagram which gives an example of the commercial practice of the Dynamit Nobel process (73). *p*-Xylene, air, and catalyst are fed continuously to the oxidation reactor where they are joined with recycle methyl *p*-toluate. Typically, the catalyst is a cobalt salt, but cobalt and manganese are also used in combination. Titanium or other expensive metallurgy is not required because bromine and acetic acid are not used. The oxidation reactor is maintained at 140–180°C and 500–800 kPa (5–8 atm). The heat of reaction is removed by vaporization of water and excess *p*-xylene; these are condensed, water is separated, and *p*-xylene is returned continuously (72,74). Cooling coils can also be used (70).

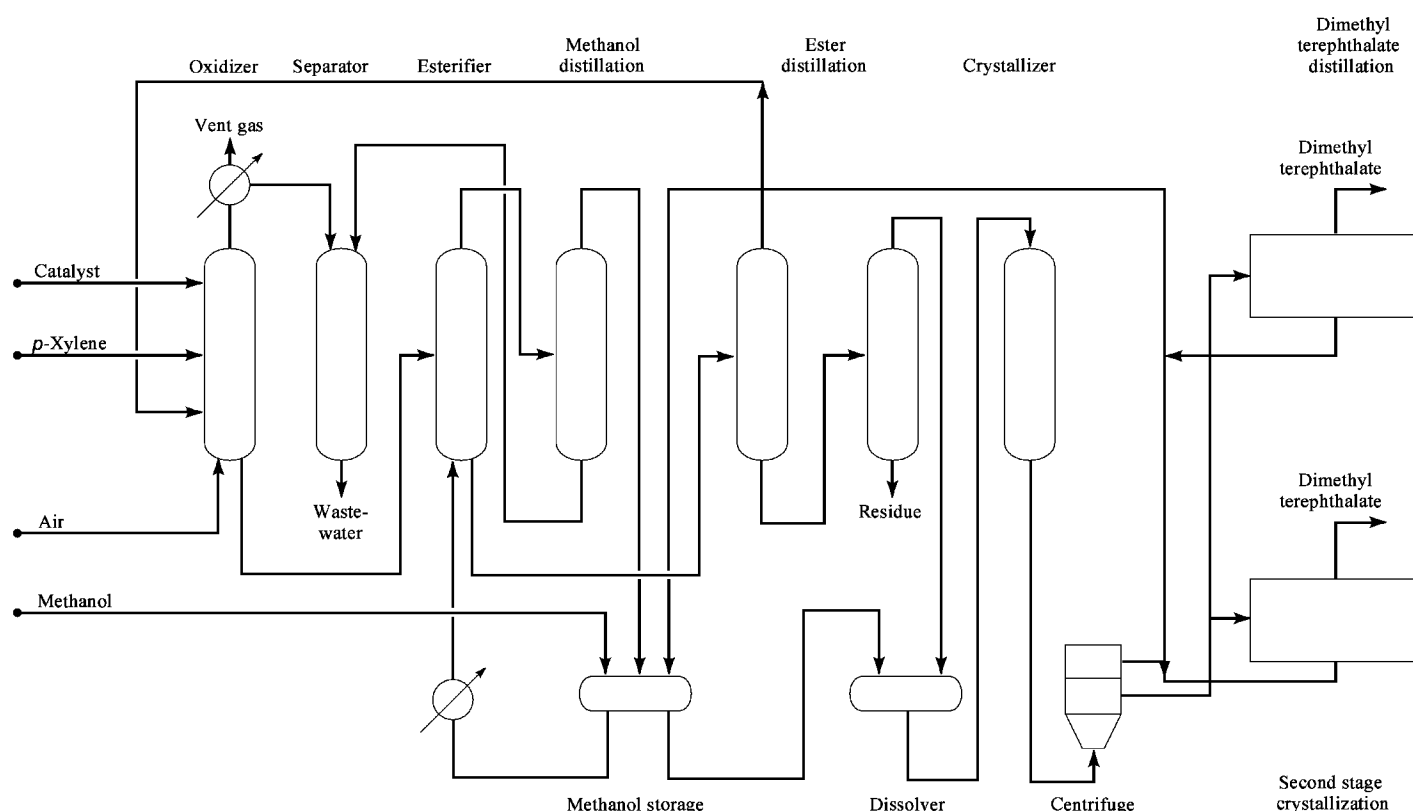


Fig. 3. The Dynamit Nobel process for dimethyl terephthalate (73).

Courtesy of Gulf Publishing Co.

The oxidation reactor effluent and methanol are sent to the esterification reactor, which operates at up to 250°C and a pressure sufficient to maintain the liquid phase. This latter is about 2500 kPa (25 atm). The oxidation products are converted to methyl *p*-toluate and dimethyl terephthalate without a catalyst. Excess methanol is supplied, and steam and vaporized methanol are removed and enter a methanol recovery column. The esterification products flow to a crude ester column, which separates the toluate from the terephthalate. The overhead stream of methyl *p*-toluate is returned to the oxidation reactor, and the bottoms stream of dimethyl terephthalate goes to a primary distillation. The distillate is dissolved in methanol, crystallized, and solid dimethyl terephthalate is recovered. The dimethyl terephthalate can then be either recrystallized or distilled to yield the highly pure material needed for the polyesterification reaction.

The overall yield of the process is at least 87 mol %, and 2.3 mol of methanol per mole of final product are needed, an excess of 15 % over the 2.0 theoretical amount. The methanol can be recycled from the manufacture of poly(ethylene terephthalate). Reported utilities consumptions per kilogram of product are 1.2 kg of 1400-kPa steam, 420 kJ of boiler fuel, and 0.5 kWh of electricity (72).

Esterification of Terephthalic Acid. Esterification of terephthalic acid is also used to produce dimethyl terephthalate commercially, although the amount made by this process has declined. Imperial Chemical Industries, Eastman Kodak, Amoco, Toray, Mitsubishi, and Mitsui Petrochemical have all developed processes. Esterification (qv) generally uses a large excess of methanol in a liquid process at 250–300°C. The reaction proceeds rapidly without a catalyst, but metal catalysts such as zinc, molybdenum, antimony, and tin can be used. Conversion to dimethyl terephthalate is limited by equilibrium, but yields of 96 % have been reported (75,76).

The crude dimethyl terephthalate is recovered and purified by distillation in most processes. Although distillation (qv) is generally a powerful separation technique, the mode of production of the terephthalic acid determines its impurity content, which in turn may make purification by distillation difficult. Processes resulting in the alteration of the impurities by catalytic treatment have been developed so that distillation can perform the necessary purification.

Polymer-Grade Terephthalic Acid. The processes described earlier for production of terephthalic acid do not yield products pure enough for use as raw materials in the production of polyester. However, polymer-grade terephthalic acid is produced through the use of a separate purification process. It is also produced by modifying the liquid-phase oxidation of *p*-xylene, and it can also be made by modifying the Henkel process. Most polymer-grade terephthalic acid made by modifying the liquid-phase oxidation process is used captively by the producing company or an affiliate, but some is sold on the open market. Polymer-grade terephthalic acid is also produced by hydrolyzing purified dimethyl terephthalate. Here too, though, the producing company also makes poly(ethylene terephthalate) and uses most of it captively.

Amoco Purification Process. The Amoco process is used to purify terephthalic acid produced by the bromine-promoted air oxidation of *p*-xylene. The main impurity in the oxidation product is 4-formylbenzoic acid [619-66-9] and the Amoco process removes this to less than 25 ppm. Metals and colored organic impurities are also almost completely removed by the purification.

A flow diagram of the continuous process is presented in Figure 4 (77,78). Crude terephthalic acid and water are fed to a mixing tank to form a slurry of at least 15 wt % terephthalic acid. The slurry is pumped to heat exchangers, which raise the slurry temperature sufficiently for the terephthalic acid to dissolve. The solution flows through a hydrogenation reactor that contains a palladium-on-carbon support catalyst. Hydrogen is added to the reactor, where it dissolves in the feed solution. Reactor temperature is held above the partial pressure of steam to maintain a liquid phase. In the reactor, 4-formylbenzoic acid is hydrogenated to *p*-toluic acid, and various colored impurities are hydrogenated to colorless products. The catalyst is highly selective; the loss of terephthalic acid by carboxylic acid reduction or ring hydrogenation is less than 1 %. The overall effect of the hydrogenation is conversion of impurities to forms which remain in the mother liquor during the subsequent crystallization step. The terephthalic acid is purified by crystallization in a series of vessels where the pressure and therefore the temperature is sequentially decreased (79). As was noted above, impurities remain in the mother liquor for the most part. The purified terephthalic acid crystals are recovered by centrifugation or filtration, followed by drying of the wet cake. Overall yield of a white, free-flowing powder is greater than 98 %.

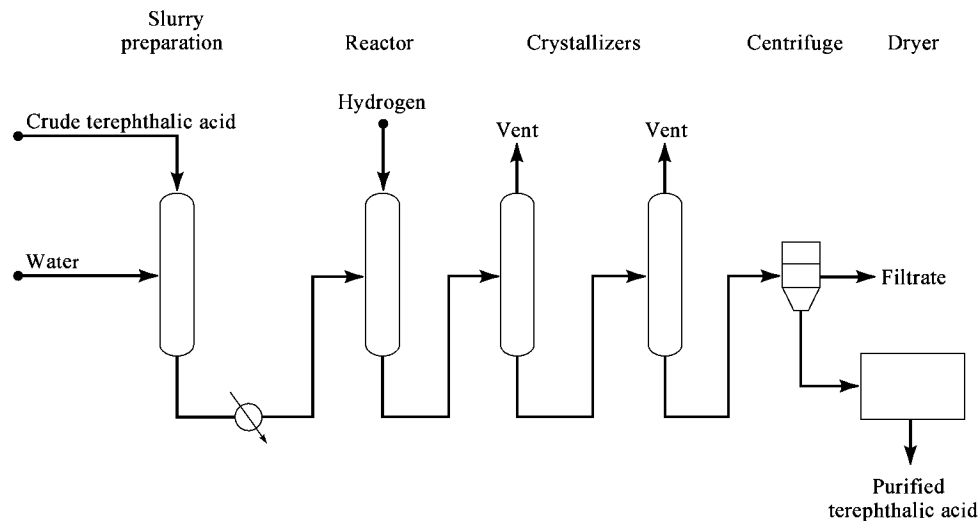


Fig. 4. The Amoco purification process for polymer-grade terephthalic acid.

Modifications of Processes Based on Air Oxidation of *p*-Xylene. Since the mid-1970s, starting in Japan, several companies have developed oxidation processes to yield relatively pure forms of terephthalic acid without a separate purification. These products, normally called medium purity terephthalic acids, contain 200–300 ppm 4-formylbenzoic acid and trace amounts of acetic acid and thus do not meet normal specifications for the highest purity grades available (80,81).

Processes for medium purity terephthalic acid have been developed by Mitsubishi, Toray, Maruzen Oil Co., Kuraray Yuka Co., Ltd., and Eastman Chemical, among others. As commercially practiced by Mitsubishi Kasei, the oxidation of *p*-xylene in acetic acid with a cobalt–manganese–bromine catalyst is performed, followed by a further oxidation in a second vessel at a higher temperature of 235–290°C. Addition of more catalyst may also be done (82). The higher temperature and longer residence times afforded by a second reactor allow digestion of crystals with consequent release of some 4-formylbenzoic acid molecules. As the 4-formylbenzoic acid enters the solution, it can be oxidized to terephthalic acid. The more intense oxidation also results in more *p*-xylene and acetic acid being converted to carbon oxides (82). After completion of the reaction, the terephthalic acid crystals are separated by centrifugation or filtration and dried.

Eastman Chemical Co. uses only cobalt and bromine, and lower temperature oxidations are held at 175–230°C (83). Solution of 4-formylbenzoic acid is obtained by using hydroclones to replace the mother liquor from the first oxidation with fresh acetic acid. A residence time of up to 2 h is used in order to allow for sufficient digestion to take place and to reduce the 4-formylbenzoic acid content to 40–270 ppm (83). Recovery of dry terephthalic acid is as described above.

Hydrolysis of Dimethyl Terephthalate. Hoechst Celanese and Formosa Chemical Fibers Corp. produce a polymer-grade terephthalic acid by hydrolysis of high purity dimethyl terephthalate. Hbbs-Troisdorf AG licenses a process with this step (70). Hydrolysis occurs at 260–280°C and 4500–5500 kPa (45–55 atm) in a hydrolysis reactor without catalysis. The overhead methanol and water vapor is separated and the methanol is returned to the dimethyl terephthalate section for reuse. The reactor liquid is crystallized, cycloned, washed, and further cooled. Finally, the slurry is centrifuged and dried. The product has less than 25 ppm of 4-formylbenzoic acid and very low levels of other impurities. There may be several hundred parts per million of monomethyl terephthalate, which is incompletely hydrolyzed dimethyl terephthalate.

Production, Storage, and Shipment. Modern terephthalic acid plants usually produce at least 250 × 10³ t annually. The relatively low selling price dictates large plants so that economies of scale can be realized, and the huge worldwide demand makes plants of this size commercially viable. Storage of terephthalic acid is in silos, and the preferred method for dimethyl terephthalate storage is molten in insulated and heated tanks. The huge-volume use of these chemicals means that bulk shipment is preferred. Rail hopper cars or hopper trucks for terephthalic acid or insulated rail tank cars or tank trucks for dimethyl terephthalate are used where possible. Further, the high purity requirements make dedicated containers highly desirable to prevent cross-contamination. One-ton bags are used for terephthalic acid and the pellet form of dimethyl terephthalate. For shipment by sea and some by truck, the containerized method is often used. The containers are either 6 or 12 m long (20 or 40 ft), and can be loaded with either 20 one-ton bags or fitted with a one-piece polyethylene liner with 20 t blown directly into the lined container.

Economic Aspects. Terephthalic acid and dimethyl terephthalate are usually sold under long-term contracts. Pricing information is at times published but actual contract prices are not revealed. Price data published in 1992 were \$0.60 /kg for terephthalic acid and \$0.57 /kg for dimethyl terephthalate (42). The price is mainly influenced by the price of *p*-xylene. The price of terephthalic acid is more than dimethyl terephthalate because a kilogram of it produces 17% more polyester. The price of dimethyl terephthalate takes this factor plus a credit for the methanol generated during polyester production into consideration.

Capacity figures for terephthalate feedstocks are given in Tables 18 and 19 (42,84,85). Healthy growth has characterized the terephthalate market since its inception after World War II, and this growth is projected to continue. Based on announced capacity expansions, most of the growth will occur for purified terephthalic acid (TA), with dimethyl terephthalate (DMT) capacity remaining essentially flat or declining as some dimethyl terephthalate plants are converted to produce medium-purity terephthalic acid. The medium-purity component will gain slightly. Terephthalic acid output and dimethyl terephthalate output were equal in 1985, but terephthalic acid will contribute over 80% of the total after the year 2000. Also seen in Table 18 is the rapid expansion of capacity in the Far East, where much of the world's garment sewing is concentrated. This expansion will continue. As in the phthalic anhydride tables, Eastern Europe–Mid-East includes areas ranging from the former Soviet Union to as far east as India.

Table 18. World Capacities^a for Terephthalic Acid/Dimethyl Terephthalate, 10³ t

Location	Year		
	1980	1988	1993
North America	3,950	4,420	4,530
Western Europe	1,740	2,270	2,550
Far East	1,460	3,200	6,210
Eastern Europe–Mid-East	600	640	1,100
South America	170	180	190
Total	7,920	10,710	14,580
TA–DMT ratio	37–63	56–44	72–28

^a Values are estimated.

Table 19. United States Capacities for Terephthalic Acid/Dimethyl Terephthalate, 10³ t

Company	Year		
	1980	1988	1992
Amoco ^a	1250	1450	1540
Du Pont ^b	900	925	880
Eastman ^b	650	670	680
Cape Industries ^b	630	635	635
<i>Total</i>	<i>3430</i>	<i>3680</i>	<i>3735</i>

^a TA only.
^b TA and DMT.

U.S. capacities, given in Table 19, show considerable stability both as to the producers and the amount produced. There have been no entrances or exits from the producer list since the 1970s. Amoco is totally a merchant supplier, whereas the other producers are poly(ethylene terephthalate) producers who exclusively or mostly satisfy their own requirements.

Amoco Chemical is the principal producer of terephthalic acid, with 18% of the world capacity in 1992. Joint ventures of Amoco add another 19%. Other significant producers of polymer-grade terephthalic acid are Imperial Chemical Industries, Mitsubishi Kasei, and Mitsui Petrochemical with 9, 8, and 5%, respectively. Several producers of poly(ethylene terephthalate) have back integrated by licensing polymer-grade terephthalic acid technology. Medium purity terephthalic acid is mainly produced by Mitsubishi Kasei and Eastman and their licensees. Dimethyl terephthalate is primarily produced by Hoechst in Europe and Hoechst-Celanese, Eastman, and Du Pont in the United States. Eastman and Du Pont have converted some dimethyl terephthalate capacity to a polymer-grade terephthalic acid. There are several small, older plants, primarily in Europe. Hoechst-Celanese and Formosa Chemical Fiber hydrolyze some dimethyl terephthalate to polymer-grade terephthalic acid.

Specifications and Standards. Polymer-grade terephthalic acid must conform to strict standards of purity because of the demanding nature of the polyesterification process. There are no industry-wide standards which can be considered international or official, but all polymer-grade terephthalic acid generally conforms to specifications and has typical analyses, as given in Table 20. There is some movement to express specifications as a target and range to ensure product consistency, as opposed to allowing, for example, any value below a certain maximum.

Table 20. Specifications for Polymer-Grade Terephthalic Acid

Property	Specification	Test method
acid number, mg KOH/g	675 ± 2	titration
ash, ppm	15 max	combustion
total significant metals, ^a ppm	9 max	atomic absorption or inductively coupled plasma
4-formylbenzoic acid, ppm	25 max	liquid chromatograph
<i>p</i> -toluic acid, ppm	125 ± 45	capillary electrophoresis
moisture, wt %	0.2 max	Karl Fischer
5% dimethylformamide soln color, APHA	10 max	colorimetry

^a Mo, Cr, Ni, Co, Fe, Ti, and Mn.

Polymer-grade terephthalic acid is over 99.9 wt % pure, exclusive of some residual water which has a specification. With this degree of purity, impurity contents are usually given instead of total purity. A purity measurement is acid number, which refers to the milligrams of potassium hydroxide needed to titrate 1 g of terephthalic acid, and is theoretically 675.5.

Equally strict purity requirements apply to dimethyl terephthalate, as shown in Table 21. Freezing point is a sensitive measure of purity and can be measured to at least 0.01°C. This is used for dimethyl terephthalate but not for terephthalic acid, which does not melt at atmospheric pressure. Because dimethyl terephthalate is preferentially handled in a hot, molten form, color stability and acid number are important. These reflect product resistance to oxidation, discoloration, and thermal degradation. Acid number is a measure of the residual unesterified carboxylic acid groups.

Table 21. Specifications for Dimethyl Terephthalate

Property	Specification	Test method
freezing point, °C	140.62 min	cryoscopy
molten color, APHA	25 max	colorimetry
color stability, 4 h at 175°C, APHA	25 max	colorimetry
aldehyde ester, ppm	30 max	gas chromatography
acid number, mg KOH/g	0.03 max	nonaqueous titration

Analytical and Test Methods. The acid number of terephthalic acid discussed above is a titration of a sample dissolved in pyridine, using a sodium or potassium hydroxide titrant. However, specifications on certain impurities are so strict that this test cannot, as a practical matter, be failed. Its use has been discontinued by some manufacturers.

Ash in terephthalic acid refers to the residue left after combustion of the sample. Ash consists of oxides of trace metals, which are determined individually by atomic absorption or inductively coupled plasma. A Karl Fischer titration is specific for the water content.

4-Formylbenzoic acid and *p*-toluic acid are determined by high performance liquid chromatography or capillary electrophoresis. Gas chromatography can be used for the aldehyde ester content of dimethyl terephthalate, this being the counterpart of 4-formylbenzoic acid in terephthalic acid.

Terephthalic acid is pure white, and molten dimethyl terephthalate is colorless. Impurities or degradation products can be yellow or brown, so the darkness of either a solution of terephthalic acid in dimethylformamide or molten dimethyl terephthalate can be compared to APHA color standards.

Health and Safety Factors. Terephthalic acid has a low order of toxicity. Inhalation by rats for 6 h/d, 5 d/wk for 4 wk produced no fatalities at a dust exposure level of 25 mg/m³. The mean acute oral toxicity for rats is over 18 g/kg (86), and for mice over 6 g/kg (87). When terephthalic acid was fed as 3% of the diet to rats, urinary calculi formed in 90 d, some of which led to cancer. High doses of terephthalic acid lead to formation of calcium terephthalate at levels exceeding its solubility in urine. This insoluble material leads to the calculi and provides a threshold below which cancer is not observed (88). Normal precautions used in handling industrial chemicals should be observed with terephthalic acid. If ventilation is inadequate, a toxic-dust respirator should be used to avoid prolonged exposure.

Dimethyl terephthalate also shows low toxicity. Inhalation by rats of dust for 4 h/d for 58 d showed no toxicological effects at levels up to 86

mg m³ (89). Mean acute oral toxicity here was reported as over 6 g kg for rats (90). Urinary calculi were also formed when dimethyl terephthalate was added at 3% to the diet of rats. As in the case of terephthalic acid, a toxic-dust respirator should be worn when ventilation is inadequate.

Molten dimethyl terephthalate burns if ignited. Dimethyl terephthalate vapor and dust, and terephthalic acid dust form explosive mixtures with air. The flammability limits for terephthalic acid and dimethyl terephthalate are given in Table 22. Dimethyl terephthalate is classified as a severe relative explosive hazard by the U.S. Bureau of Mines. Relative to phthalic anhydride, dimethyl terephthalate has a lower ignition sensitivity and a higher explosive severity (41,42,91). Fires can be extinguished with dry chemical, carbon dioxide, water or water fog, or foam. When dimethyl terephthalate is near flames, it may melt, flow, and contribute to the spreading of the fire.

Table 22. Flammability and Explosivity of Terephthalic Acid and Dimethyl Terephthalate

Property	Terephthalic acid	Dimethyl terephthalate
	<i>Dust cloud</i>	
explosibility index	6.9	>10
ignition sensitivity	3	5.9
ignition temperature, °C	680	570
minimum igniting energy, J ^a	0.020	0.020
minimum explosive concentration, g L	0.050	0.030
relative flammability, % inert		
by furnace ignition		>90
by spark ignition	80	>90
limiting oxygen concentration, %		
by furnace ignition at 850°C		12
by spark ignition	15	6
explosive severity	2.3	5.8
maximum pressure at concentration L, kPa ^b		
0.1 g L	227	372
0.5 g L	503	634
1.0 g L	579	641
2.0 g L	531	724
average pressure rise at concentration, g L, 10 ³ kPa s ^b		
0.1 g L	3.4	15
0.5 g L	12	21
1.0 g L	9.6	13
2.0 g L	4.1	11
maximum pressure rise at concentration, g L, 10 ³ kPa s ^b		
0.1 g L	14	62
0.5 g L	41	83
1.0 g L	26	41
2.0 g L	8.3	34
	<i>Molten liquid</i>	
flash point, Cleveland open cup, °C		146–147
fire point, Cleveland open cup, °C		155

^a To convert J to cal, divide by 4.184.

^b To convert kPa to psi, multiply by 0.145.

Uses. Essentially all polymer-grade terephthalic acid and dimethyl terephthalate are used to make saturated polyesters, the great majority being poly(ethylene terephthalate). Poly(ethylene terephthalate) is employed to make fiber and is the largest-volume synthetic fiber in the world. Fiber use makes up 73% of worldwide poly(ethylene terephthalate) production (92). It is used for woven and knitted fabrics for clothing, draperies, upholstery, and carpeting. In clothing applications, it is usually blended with other fibers, primarily cotton (qv). Applications for high strength polyester continuous fibers are reinforcing cord for tires, V-belts, conveyor belts, and hoses (see FIBERS, POLYESTER).

Polyester film consumes 7% of production. When coated with a chemical emulsion, it is used as x-ray and microfilm; when coated with a magnetic emulsion, it is used for audio and video tapes; and when coated with an adhesive, it is used for wrapping and sealing tapes (see FILMS AND SHEETING).

The fastest growing application for poly(ethylene terephthalate) is in packaging, especially bottles. It currently accounts for 15% of production. U.S. FDA approval of polyester bottles has made it the polymer of choice for beverage and food applications. Polyester bottles have almost displaced glass carbonated beverage bottles due to their clarity, light weight, and shatter-resistant properties. Applications in food packaging are expanding rapidly for the same reasons (see POLYESTERS, THERMOPLASTIC).

Small amounts of polymer-grade terephthalic acid and dimethyl terephthalate are used as polymer raw materials for a variety of applications, eg, adhesives and coatings. They are also used to make high performance polymers or engineering resins. Poly(ethylene terephthalate) is itself an engineering resin, although one more widely used is poly(butylene) terephthalate, formed by reaction with 1,4-butanediol as the comonomer. 1,4-Cyclohexanedimethanol can be the comonomer. Terephthalic acid can also be the diacid in specialty nylons (see POLYAMIDES). Specialty fibers, including certain high modulus aramid fibers, are made from terephthalic acid–dimethyl terephthalate derivatives.

Derivatives. In general, the esters of terephthalic acid derived from saturated alcohols undergo the same reactions as dimethyl terephthalate. Some physical properties of six of these esters are listed in Table 23. The di-*n*-butyl and di-2-ethylhexyl esters find use as plasticizers (qv). Terephthaloyl chloride, which is prepared by reaction of terephthalic acid and thionyl chloride, is used to prepare derivatives of terephthalic acid.

Table 23. Properties of terephthalic acid ester and Terephthaloyl Chloride

Ester	CAS Registry Number	Mp, °C	Bp, °C (Pa) ^a	Specific gravity °C
diethyl terephthalic acid esters	[636-09-9]	43–44	140–142 400 Pa	1.102
di- <i>n</i> -propyl tereph-thalic acid esters	[1962-74-9]	25	164–166 530 Pa	1.072
di- <i>n</i> -butyl tereph-thalic acid esters	[1962-75-0]	16.6–18.1	181–189 270 Pa	1.045
di(2-ethylhexyl) terephthalic acid esters	[6422-86-2]	63.5	186–192 13 Pa	0.982
diisooctyl terephthalic acid esters	[4654-26-6]	amorphous solid	189–190 6.7 Pa	0.981
diallyl terephthalic acid esters	[1026-92-2]	10	140–143 130 Pa	1.138
terephthaloyl chloride terephthalic acid	[100-20-9]	78	259	

esters
^a To convert Pa to mm Hg, divide by 133.3.

Isophthalic Acid

Like terephthalic acid, isophthalic acid is used as a raw material in the production of polyesters. Much of the isophthalic acid is used for unsaturated polyesters, whereas terephthalic acid is used almost exclusively in saturated (thermoplastic) polyesters. However, a considerable amount of isophthalic acid is used as a minor comonomer in saturated polyesters, where the principal diacid is terephthalic acid. The production volume of isophthalic acid is less than 2° o that of terephthalic. Isophthalic acid was formerly produced in technical or crude grades and only a small amount was purified. Now, however, it is all purified to a standard similar to that of terephthalic acid.

Physical and Chemical Properties. Selected physical and chemical properties of isophthalic acid are shown in Tables 24 and 25.

Table 24. Physical Constants and Properties of Isophthalic Acid

Property	Value
mp (closed tube), °C	345–348
vapor pressure, kPa ^a	
at 100°C	0.009
125°C	0.08
230°C	0.23
260°C	1.03
290°C	3.98
specific gravity at 4°C	1.53
heat of combustion at 25°C, kJ mol ^b	3202
heat of formation at 25°C, kJ mol ^b	802
heat of sublimation at 25°C, kJ mol ^b	106.7

^a To convert kPa to mm Hg, multiply by 7.5.

^b To convert J to cal, divide by 4.184.

Table 25. Solubilities of Isophthalic Acid^a

Solvent	Temperature, °C				
	25°C	50°C	100°C	150°C	200°C
water	0.012	0.035	0.32	2.8	25
acetic acid (glacial)	0.23	0.41	1.3	4.3	11.1
methanol	2.5	4.0			
1-propanol	1.7	2.7	7.0		
dimethylformamide	37				
dimethyl sulfoxide	64				

^a g 100 g solvent.

Manufacture and Processing. Isophthalic acid is synthesized commercially by the liquid-phase oxidation of *m*-xylene [108-38-3]. The chemistry of the oxidation is almost identical to that of *p*-xylene oxidation to terephthalic acid, and production facilities can be used interchangeably for these two dicarboxylic acids. However, because isophthalic acid is more soluble than terephthalic acid in reaction solvents as can be seen by comparing data in Tables 16 and 25, crystallization equipment is more important in isophthalic acid facilities.

Vapor phase oxidation of *m*-xylene is not satisfactory. Its vapor pressure, although low, is higher than that of terephthalic acid. However, isophthalic acid remains on the catalyst until it decomposes, resulting in poor or no yield. Summaries of patent literature describing oxidation of *m*-xylene with oxygen-containing gases, with nitric acid, with sulfur compounds, and by other routes are given in Reference 93.

Chevron Chemical Co. began commercial production of isophthalic acid in 1956. The sulfur-based oxidation of *m*-xylene in aqueous ammonia at about 320°C and 7,000–14,000 kPa produced the amide. This amide was then hydrolyzed with sulfuric acid to produce isophthalic acid at about 98° o purity. Arco Chemical Co. began production in 1970 using air oxidation in acetic acid catalyzed by a cobalt salt and promoted by acetaldehyde at 100–150°C and 1400–2800 kPa (14–28 atm). The crude isophthalic acid was dissolved and recrystallized to yield a product exceeding 99° o purity. The Arco technology was not competitive and the plant was shut down in 1974.

Amoco Chemical Co. used the cobalt–manganese–bromine catalyst system and, in 1958, started isophthalic acid production by oxidizing mixed xylenes and then separating the acids. Acid separation proved difficult, and mixed xylenes were not necessarily in proportion to the market demand for the acids. Feedstock to the oxidation was then switched so that the acid product had the same isomeric content as the xylene feed. Today, only pure *m*-xylene is fed. Formerly, an 85° o:15° o *m*-xylene–*p*-xylene mix was also used to make a mixed acid of the same proportion, which was sold and used commercially. Figure 2 and the description of the terephthalic acid process also represent the *m*-xylene-to-isophthalic acid process. The surge vessel in Figure 2 is a crystallizer as well as pressure let-down vessel in isophthalic acid production. The reaction is carried out at 170–230°C and 2000–3500 kPa (20–35 atm). As with terephthalic acid, the temperature–pressure relationship is determined by the boiling of the acetic acid–water solvent. This vaporization with reflux of the condensed solvent removes the heat of oxidation. The solubility of isophthalic acid gives a solution or a dilute slurry in the reactor depending on the temperature. The isophthalic acid must therefore be crystallized downstream.

Impurities in isophthalic acid from the oxidation process are analogous to those in terephthalic acid, eg, 3-formylbenzoic acid and *m*-toluic acid. Also present are other impurities such as benzoic acid and residual catalyst metals. All isophthalic acid made by this liquid-phase oxidation is now purified in a process similar to that used for terephthalic acid, as shown in Figure 4. Lower temperatures are used owing to the greater solubility of isophthalic acid vs terephthalic acid.

Societa Italiana Serie Acetica Sintetica (SISAS) produces isophthalic acid commercially by a proprietary process (94,95). They have installed purification facilities for hydrogenation and crystallization similar to those used for terephthalic acid.

Production, Storage, and Shipment. A plant of 50,000-t annual capacity could be considered viable for production of isophthalic acid, although no new plants have been constructed since the early 1980s. Storage of isophthalic acid is in silos. Shipment is in 22.7- and 25-kg bags, 0.5-t and 1-t bags, or hopper trucks. The far lower production quantity of isophthalic acid and its more varied applications vs terephthalic acid mean that high volume

rail hopper cars are not used to any extent, and hopper truck shipment is also small.

Economic Aspects. Isophthalic acid in North America sold for \$1.19–\$1.32 kg in 1994, depending on the shipment method. The price of *m*-xylene plays a role, although not to the same extent as *p*-xylene in terephthalic acid. The far lower production volumes and smaller plant sizes for isophthalic acid do not give the same economies of scale.

Worldwide capacity available for production of isophthalic acid was about 270 × 10³ t in 1994. About 200 × 10³ t was actually produced. Amoco Chemical with plants in the United States and Europe is the principal producer, with over 60^o % of the production, and AGIC, an Amoco–Mitsubishi Gas Chemical joint venture in Japan, and SISAS produce the remainder.

Specifications and Standards. The specifications for isophthalic acid used by Amoco are given in Table 26. Because the same process is used, the specifications closely parallel those of terephthalic acid. These specifications are more demanding than those used before 1990, when a technical-grade isophthalic acid as produced by the oxidation process was available. Owing to the introduction of a water-based purification process including a recrystallization, impurity limits are much tighter. Purified isophthalic acid is over 99.9 wt ^o % pure, exclusive of some residual water which has a specification. With this degree of purity, impurity contents are usually given instead of total purity.

Table 26. Specifications for Isophthalic Acid

Property	Specification	Test method
isophthalic acid, wt ^o %	99.8 min	polarography
ash, ppm	18 max	combustion
metals, ^a ppm	1–3 max	atomic absorption or inductively coupled plasma
3-formylbenzoic acid, ppm	25 max	polarography
<i>m</i> -toluic acid, ppm	150 max	esterification gas chromatography
moisture, wt ^o %	0.1 max	Karl Fischer

^a Al, Ca, Cr, Co, Fe, Mn, Mo, Ni, K, Na, and Ti.

Analytical and Test Methods. Ash in isophthalic acid refers to the residue left after combustion of the sample. Ash consists of oxides of trace metals that are determined individually by atomic absorption or inductively coupled plasma. A Karl Fischer titration is specific for the water content.

3-Formylbenzoic acid and *p*-toluic acid can be determined by high performance liquid chromatography. In some cases, polarography is used for 3-formylbenzoic acid and esterification gas chromatography for the *m*-toluic acid content.

Isophthalic acid is pure white, but some impurities can be yellow. A measure of these impurities is obtained by the yellowness, or b-value, of the sample.

Health and Safety Factors. Isophthalic acid has a low order of toxicity. Inhalation by rats for 4 h at 11.4 g m³ showed no toxicity. The LD₅₀ level for rats is high at 10.4 g kg (96). As with terephthalic acid, isophthalic acid was found to form urinary tract calculi in rats in 90 d when it constituted 3^o % of their diet. This led to some cancer owing to the presence of the calculi. Some mild eye irritation is possible, so eye protection should be worn. Otherwise, normal precautions used in handling industrial chemicals should be observed with isophthalic acid.

Isophthalic acid dust forms explosive mixtures with air at certain concentrations. These concentrations and other information on burning and explosiveness of isophthalic acid dust clouds are given in Table 27 (40,41). Fires can be extinguished with dry chemical, carbon dioxide, water or water fog, or foam.

Table 27. Flammability and Explosivity of Isophthalic Acid

Property	Value
explosibility index	4.0
ignition sensitivity	3.3
ignition temperature, °C	700
minimum igniting energy, J ^a	0.025
minimum explosive concentration, g L	0.035
limiting oxygen concentration (spark ignition), ^o %	14

^a To convert J to cal, divide by 4.184.

Uses. About 35^o % of the isophthalic acid is used to prepare unsaturated polyester resins. These are condensation products of isophthalic acid, an unsaturated dibasic acid, most likely maleic anhydride, and a glycol such as propylene glycol. The polymer is dissolved in an inhibited vinyl monomer, usually styrene with a quinone inhibitor. When this viscous liquid is treated with a catalyst, heat or free-radical initiation causes cross-linking and solidification. A range of properties is possible depending on the reactants used and their ratios (97).

Because of higher raw material and manufacturing costs, isopolyesters are more expensive than phthalic anhydride-based polyesters. However, the isopolyesters have higher molecular weights; give improved water, chemical, and weathering resistance; and have higher heat-distortion temperatures. Most isopolyesters are used with glass fiber reinforcement. These include sheet molding compounds, bulk molding compounds, and filament windings. Uses for reinforced isopolyesters include storage tanks and piping, architectural panels for exterior or interior use, automobile body panels, swimming pools and spas, and boat hulls, including some over 45-m long. In particular, isopolyesters show much less blistering in boat hulls, owing to their greater hydrolytic stability. Applications of isopolyesters without reinforcement include gel coats, concrete coatings, and terrazzo glaze coats.

The second application of isophthalic acid, with about 30^o % of the output, is for alkyd coatings. Isophthalic alkyds are made by reaction of isophthalic acid and other polybasic acids such as trimellitic anhydride, polyhydric alcohols such as glycerol, and usually fatty oils or acids such as tall oil. These resins can be formulated into solvent-based, water-based, or solventless coatings for both commercial and consumer applications. As with isopolyesters, these coatings are more expensive than those made with phthalic anhydride, but they offer faster drying; higher hardness; and better resistance to weathering, marring, and high temperatures.

The third, and fastest growing, area of isophthalic acid use is in other types of polymers, primarily as a minor comonomer with terephthalic acid in saturated polyesters. Over 20^o % of the isophthalic acid is sold in this application. One rapidly expanding use is in polyester beverage bottles where addition of up to 3^o % isophthalic acid to the terephthalic acid allows faster production of more complex shapes. In this way, single piece bottles can be made, vs a round-bottomed bottle that needs a separate base cup. Fibers are also modified with isophthalic acid.

Isophthalic acid is also used in formulations for adhesives, inks (qv), wire enamels, and dental materials (qv). Copper isophthalate [10027-31-3] is an ingredient in algicides and fungicides (98).

Derivatives. Commercially significant derivatives of isophthalic acid are its diesters with saturated alcohols and isophthaloyl chloride. This derivation is similar to that of terephthalic acid. Some properties of these compounds are given in Table 28. Plasticizers form the main use of the diesters, these being the dimethyl, dioctyl, and di(2-ethylhexyl) isophthalates. Diallyl isophthalate is a cross-linking agent for high temperature resistant polybenzimidazoles. Isophthaloyl chloride is used in the manufacture of high temperature-resistant polyamide fibers, films, dyes, and protective coatings.

Table 28. Properties of Isophthalic Esters and Isophthaloyl Chloride

Ester	CAS Registry Number	Mp, °C	Bp, °C (Pa) ^a	Specific gravity
dimethyl isophthlate	[1459-93-4]	66–68	155–159 1333 Pa	1.194
diethyl isophthlate	[636-53-3]	11.5	302	1.124
diisooctyl isophthlate	[137-89-3]		127–138 53 Pa	0.987
di(2-ethylhexyl) isophthlate	[137-89-3]		241 666 Pa	0.984
diallyl isophthlate	[1087-21-4]		150–152 120 Pa	0.114
diphenyl isophthlate	[744-45-6]	134–138		
isophthaloyl chloride isophthlate	[99-63-8]	43–44	276	1.388

^a To convert Pa to mm Hg, divide by 133.3.

Trimellitic Acid and Trimellitic Anhydride

Of the three benzenetricarboxylic acids, only trimellitic acid as the anhydride is commercially produced in large volume, by liquid-phase air oxidation of either pseudocumene or dimethyl benzaldehyde. The pseudocumene oxidation is another variant of the cobalt–manganese–bromine catalyst in acetic acid solvent as described in the terephthalic acid section. The acid is available as a laboratory chemical (99). The IUPAC name of trimellitic anhydride is 5-isobenzofurancarboxylic acid (1,3-dihydro-1,3-dioxo).

Owing to the dual functional groups of acid and anhydride, trimellitic anhydride imparts performance enhancements to its end-use applications, many of which are similar to those of phthalic anhydride. In many cases, products made from trimellitic anhydride exhibit properties superior to those of phthalic anhydride. For example, trimellitate esters are used as plasticizers for poly(vinyl chloride) where higher permanency is required, especially at high temperatures. Accordingly, they are used for high temperature wire insulation and applications such as automobile interiors which become heated during summer months. Other important uses of trimellitic anhydride are solvent-borne and water-borne coatings, powder coatings, wire enamels used in electric motor and generator windings, alkyd resins, amide–imide polymers, and epoxy curing.

Physical and Chemical Properties. Trimellitic acid and trimellitic anhydride are odorless white crystalline solids in their pure form. The acid is reasonably stable up to the melting point, where dehydration to the anhydride occurs. The anhydride reacts with atmopsheric moisture, even at room temperature, to revert to the acid. Physical properties of the acid and its anhydride are listed in Tables 29–31.

Table 29. Physical Constants of Trimellitic Acid and Trimellitic Anhydride

Property	Trimellitic acid	Trimellitic anhydride
mp, °C	238 dec	168
bp, °C		390
critical temperature, °C		569
critical pressure, kPa ^a		2968
heat of vaporization, kJ mol		
at 200°C		106.2
300°C		100.4
specific heat, J (kg·K) at 204°C ^b		1966
flash point, °C		227
heat of formation at 25°C, kJ mol ^b	1179	
heat of fusion at 168°C, kJ mol ^b		29.04
acid number, mg KOH g	801	876.1
relative vapor density (air = 1)		6.6
pH of saturated aqueous solution	2.0	

^a To convert kPa to psi, multiply by 0.145.

^b To convert J to cal, divide by 4.184.

Table 30. Physical Properties of Liquid Trimellitic Anhydride

Temperature, °C	Vapor pressure, kPa ^a	Density, g cm ³	Viscosity, mPa·s(c·P)	Specific heat, J (kg·K) ^b
180°C		1.349	11.8	1930
200°C	0.3	1.332	9.1	1958
220°C		1.314	7.3	
240°C	1.4	1.296	5.8	2021
260°C	2.9	1.278		
300°C	10.7			
350°C	41			
390°C	101			

^a To convert kPa to mm Hg, multiply by 7.5.

^b To convert J (kg·K)) to cal (g·°C), divide by 4184.

Table 31. Solubilities of Trimellitic Acid and Trimellitic Anhydride at 25°C^a

Solvent	Trimellitic acid	Trimellitic anhydride
water	2.1	reacts

acetic acid (glacial)	1	reacts
acetone	7.9	49.6
2-butanone		36.5
dimethylformamide	31.3	15.5
ethanol (absolute)	25.3	reacts
ethyl acetate	1.7	21.6
cyclohexanone		38.4
mineral spirits	0.03	0.06
mixed xylenes	0.006	0.4

^a g 100 g solvent.

Manufacture and Processing. Trimellitic anhydride has been commercially available only since the early 1960s. In theory, trimellitic acid could be obtained by oxidizing almost any 1,2,4-tri-substituted alkylbenzenes with a suitable oxidizing agent. Either nitric acid oxidation of pseudocumene (1,2,4-trimethylbenzene) [95-63-6] or a combination of liquid-phase air and nitric acid oxidation had been commercially practiced in Japan and in the Federal Republic of Germany in the 1960s. Those operations were discontinued in the early 1970s. A poor yield and the need for extensive purification to remove by-products made the processes uneconomical. Unlike vapor-phase oxidation of *o*-xylene to phthalic anhydride, vapor-phase oxidation of pseudocumene results in a poor trimellitic anhydride yield. A significant product from the vapor-phase oxidation is phthalic anhydride, as a result of an extensive decarboxylation.

Amoco Chemical Co. is the sole U.S. producer of trimellitic anhydride, with a 47,500-t/yr plant in Joliet, Illinois. The Amoco process is the liquid-phase air oxidation of pseudocumene using acetic acid as a solvent and a cobalt–manganese–bromine catalyst system (100–103). The oxidation process is much like that for terephthalic acid or isophthalic acid and is illustrated in Figure 5. Trimellitic acid is recovered from the oxidation reactor effluent by a suitable means, and the acetic acid solvent is sent to the solvent tower to remove the water of reaction and for recycling. Trimellitic acid is thermally dehydrated to form trimellitic anhydride and further purified by fractional distillation to obtain high purity trimellitic anhydride.

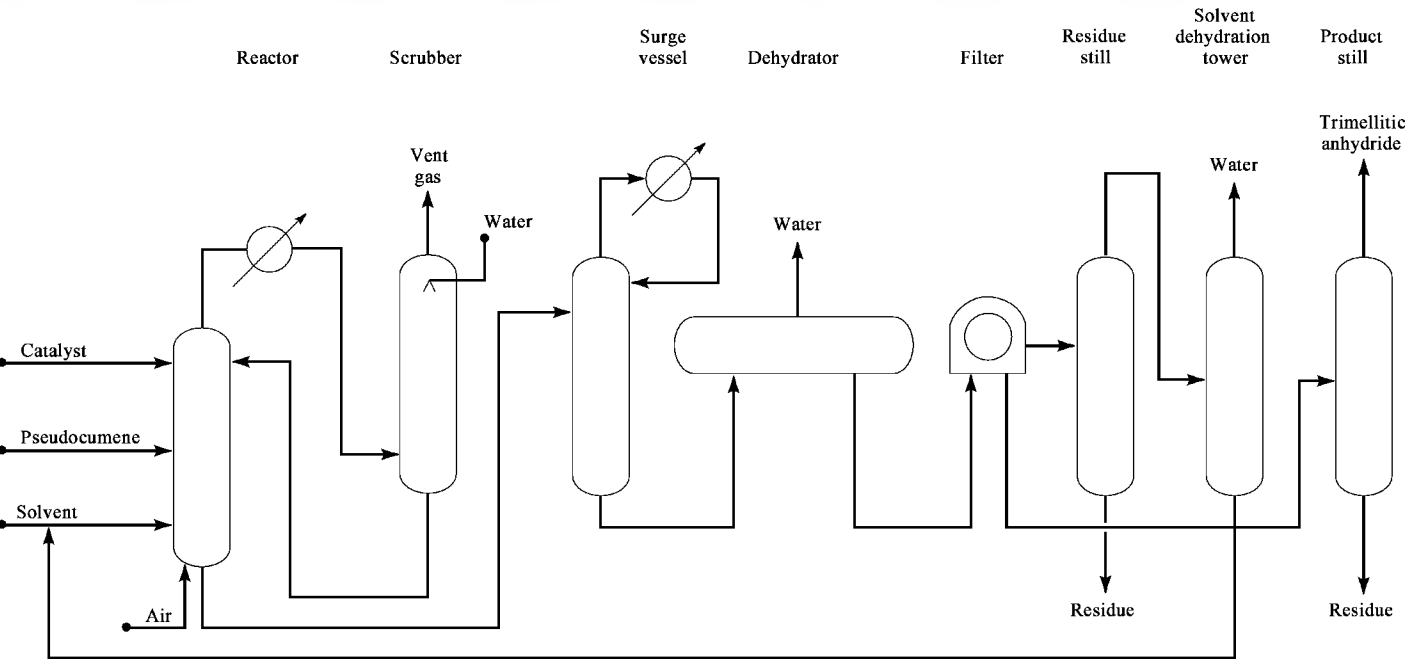


Fig. 5. The Amoco process for trimellitic anhydride.

Mitsubishi Gas Chemical Co. of Japan started the production of trimellitic anhydride at their Mizushima Works in 1986 with a name plate capacity of 15×10^3 t/yr (104,105). Pyromellitic anhydride is also produced from the same unit on a blocked out basis. Hence, the actual production of trimellitic anhydride is expected to be less than the stated capacity. The Mitsubishi Gas Chemical process for the manufacture of trimellitic anhydride starting from *m*-xylene is shown in Figure 6. *m*-Xylene is first carbonylated with carbon monoxide in the presence of boron trifluoride and hydrogen fluoride to form 2,4-dimethylbenzaldehyde. The 2,4-dimethylbenzaldehyde [15764-16-6] is decomplexed from the acids, purified, and oxidized to trimellitic acid. The air oxidation is carried out in the liquid phase using water as a solvent and manganese bromide as a catalyst along with hydrobromic acid promoter. Trimellitic acid is recovered from the reactor effluent and is subjected to normal dehydration and purification steps to obtain high quality trimellitic anhydride (106–108).

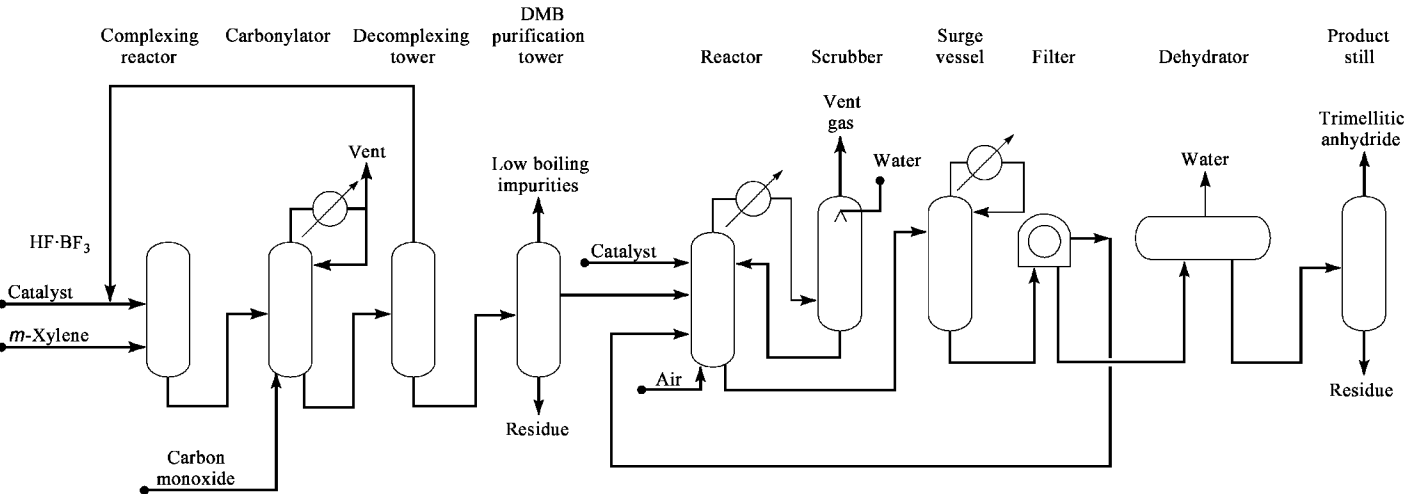


Fig. 6. The Mitsubishi Gas Chemical process for trimellitic anhydride. DMB = dimethylbenzaldehyde.

In 1991, Alusuisse Italia announced the construction of a 20×10^3 t yr trimellitic anhydride unit at Bergamo, Italy (109) and commenced production in late 1994. A European patent application (110) assigned to Alusuisse suggests that the process used is very similar to that of Amoco, that is, pseudocumene is air oxidized in the liquid phase using heavy-metal catalysts and bromine. A number of other companies have shown interest in the production of trimellitic anhydride (111–113).

Production, Storage, and Shipment. Trimellitic anhydride is produced by Amoco, Mitsubishi Gas Chemical, and Alusuisse as described above. It is available only as flakes (114). It is packaged in 906- and 500-kg bulk fabric bags and 22.6-kg kraft bags for the U.S. domestic use, and in 25-kg kraft bags for export. All bags contain moisture-barrier liners to protect the anhydride from water during storage and shipping.

Economic Aspects. When trimellitic anhydride was introduced in semicommercial quantities in 1962, it was priced at \$1.19 kg. The price was reduced to \$0.55 kg as it became available in commercial quantities in 1968. The mid-1994 price was quoted as \$2.31 kg from Amoco, fob, the Joliet, Illinois plant. A price history is given in Table 32. Although trimellitic anhydride production and sales figures are not available, the published U.S. Tariff Commission's production data for trimellitate esters provides data for the trimellitic anhydride demand trend in the United States, since the largest single use of trimellitic anhydride is for the trimellitate esters (115). These data are given in Table 32.

Table 32. U.S. Trimellitic Anhydride Prices and Trimellitate Ester Production

Year	Trimellitic anhydride price, \$ kg	Trimellitate ester production, 10 ³ t yr
1975	0.77	7.3 ^a
1980	1.47	
1985	1.87	22.1
1990	2.09	23.0
1993	2.31	24.0

^a Data not available.

Specification, Standards, and Analytical and Test Methods. The sales specifications of trimellitic anhydride from Amoco Chemical Co., the sole U.S. producer, are shown in Table 33. Trimellitic anhydride is a trifunctional, highly reactive aromatic anhydride. The reactivity of the anhydride function is important in many applications in adhesives, coatings, and amide–imide polymers, among many others. Hence, determination of the anhydride content is essential (116). A preferred method and one subject to the least interference from contaminants is achieved by comparing the results of methanolysis and hydrolysis of trimellitic anhydride. In the first step, the anhydride is hydrolyzed with hot water and titrated with standard base. The titration provides a measure of both carboxylic groups in the anhydride configuration, of the free-acid group, and of any free acid in the sample. In the second step, the anhydride is esterified to its half ester with excess absolute methanol and is titrated. This titration provides a measurement of one carboxyl group from the anhydride function, of the free carboxyl group, and of any free acid in the sample. The anhydride content is calculated from the difference between the two titrations.

Table 33. Specifications for Trimellitic Anhydride

Property	Specification	Test method
physical appearance	white flakes	visual
freezing point, °C	166 min	cryoscopy
anhydride content, wt %	95.5 min	methanolysis
color, ΔE	2.4 max	total color difference

The freezing point of trimellitic anhydride, the maximum temperature reached during crystallization of a molten sample, is a measure of the product purity. Impurities and trimellitic acid formed by hydrolysis depress the freezing point.

A measure of the color developed by impurities when trimellitate esters are produced can be correlated with the anhydride color measurement. The method measures the color difference in light transmittance between a trimellitic solution and a 3.0 N sodium hydroxide solution as a reference. The difference in light transmittance or ΔE (total color difference) is obtained using a colorimeter.

Health and Safety Factors. Trimellitic anhydride may cause respiratory irritation and, in some cases, individuals exposed over long periods may become sensitized and experience mild to severe reactions upon subsequent exposure. It should be handled with caution and treated as a toxic agent in the workplace because exposure may result in irritation of the pulmonary tract, eyes, nose, and skin (117), immunological sensitization and, in rare cases, hemolytic anemia and noncardiac pulmonary edema. Allowable and recommended exposure limits have been established by the Occupational Safety and Health Administration for a permissible exposure limit (PEL), the American Conference of Governmental Industrial Hygienists for a threshold limit value (TLV), and Amoco for a ceiling limit are all 0.4 mg m³. The PEL and TLV are an 8-h time-weighted average. The mean lethal acute oral dosage in rats is 5.6 g kg. Handling precautions include effective ventilation and use of respirators, protective clothing, and goggles when exposure to dust is expected.

Hazardous Material Identification System National Fire Protection Association (HMIS NFPA) codes for trimellitic anhydride are health, 3; flammability, 1; and reactivity, 1. Flaked or molten trimellitic anhydride will burn if ignited. High dust concentrations in the air from either trimellitic anhydride or acid have a potential for combustion or explosion. High voltage static electricity buildup is possible when handling trimellitic anhydride; therefore, adequate precautions, including bonding and grounding of equipment, as well as use of inert gas purge, should be observed. The vapor from molten trimellitic anhydride forms explosive mixtures with air. Estimated upper flammable limit for trimellitic anhydride in air is 7 vol %, and lower limit, 1 vol %. Other health- and safety-related properties are flash point, 227°C (ASTM D1310); and minimum explosive dust concentration in air, 35 g m³ (117,118).

Uses. The largest end-use application of trimellitic anhydride is as high performance poly(vinyl chloride) plasticizers in the form of triesters of aliphatic alcohols. The second most important use is in coatings applications; conventional solvent-borne, water-borne, and powder coatings. The third largest market is the use in wire enamels for high temperature performance. These three end uses make up about 95% of total trimellitic anhydride usage. Other minor applications are as epoxy curing agent (119), textile sizing agent (120–122), rubber curing accelerator (123), electrostatic toner binder (124,125), and a vinyl cross-linking agent as triallyl trimellitate (126).

The trimellitates used as plasticizers for poly(vinyl chloride) are higher molecular weight triesters made with 3 mol of monohydric aliphatic alcohols and 1 mol of trimellitic anhydride. The plasticizer-grade alcohols typically include the range of 7–13 carbon atoms. The trimellitates are approximately 40% higher in molecular weight than the corresponding phthalates owing to the third ester group presence. Accordingly, the trimellitates are significantly less volatile and less water-soluble than the corresponding phthalates, allowing the trimellitates to be used in a variety of special applications requiring high temperature permanency, low temperature flexibility, and low water extractability. A wire and cable insulation capable of withstanding 90–105°C environment, launderable vinyl apparel, refrigerator and freezer gaskets, antifogging automobile upholstery and trim, and the manufacture of floor covering are a few examples. Estimated production of trimellitates by individual type in 1990 (115) is as follows (10³ t yr): tri(2-ethylhexyl) trimellitate (TOTM), 10.9; triisononyl trimellitate (TINTM), 10.0; linear C₇–C₉ and C₈–C₁₀ trimellitates, 1.8; and tri(*n*-hexyl) and other trimellitates, 0.3.

Growing concerns over solvent costs and atmospheric pollution from solvent-borne coatings necessitate low volatile organic compounds (VOC)

formulations, with an ultimate target of zero VOC coatings. The trifunctionality of trimellitic anhydride permits synthesis of many polymers useful in formulating paints and coatings for water-based and conventional solvent-based resins. Particularly interesting are the water-soluble resins that use the third functional group as a solubilizing site. Through a proper selection of polymer ingredients and cross-linking agents, trimellitic anhydride-based resins can be formulated into air-dry, oven-cured, or powder coatings with a wide range of application and performance properties, while minimizing environmental concerns (127–132).

Polymers based on trimellitic anhydride are widely used in premium electromagnetic wire enamels requiring high temperature performance. Several types of trimellitic anhydride-derived polymers are used as wire enamels: poly(amide–imide)s (133), poly(ester–imide)s (134), and poly(amide–imide–ester)s (135). Excellent performance characteristics are imparted by trimellitic anhydride-based polymers for wire enamel requirements of flexibility, snap, burnout, scrap resistance, heat shock, and dielectric strength.

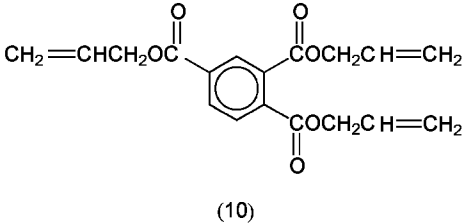
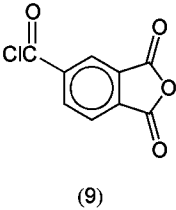
Derivatives. The dual functionality of trimellitic anhydride makes it possible to react either the anhydride group, the acid group, or both. Derivatives of trimellitic anhydride include ester, acid esters, acid chloride, amides, and amide–imides (136). Trimellitate esters are the most important derivatives, and physical properties of more significant esters are listed in Table 34.

Table 34. Properties of Trimellitic Acid Esters and the Chlorocarbonyl Anhydride

Ester	CAS Registry Number	Bp, °C (Pa) ^a	Specific gravity
trimethyl trimellitate	[2459-10-1]	194 1600 Pa	
triethyl trimellitate	[14230-18-3]	230 3730 Pa	
tri- <i>n</i> -butyl trimellitate	[1726-23-4]	206 133 Pa	1.059
tri- <i>n</i> -hexyl trimellitate	[1528-49-0]	243 533 Pa	1.006
triisooctyl (TIOTM) trimellitate	[27251-75-8]	283 ^b 333 Pa	0.989
tri(2-ethylhexyl) (TOTM) trimellitate	[3319-31-1]	260 ^b 133 Pa	0.989
triisononyl (TINTM) trimellitate	[53894-23-8]		0.979
triisodecyl trimellitate	[36631-30-8]	272 133 Pa	0.969
triallyl trimellitate	[2694-54-4]	109–116 600 Pa	
4-chlorocarbonyl-trimellitic 1,2-anhydride	[1204-28-0]	175 ^c 2000 Pa	

^a To convert Pa to mm Hg, divide by 133.3.
^b Melting point < 40° C.
^c Melting point 68–69° C.

4-Chlorocarbonyltrimellitic acid 1,2-anhydride [1204-28-0] (9), is used in the preparation of esters and amide–imide polymers. Triallyl trimellitate [2694-54-4] (10) is used as a cross-linking or co-curing agent for ethylene-derived rubbers and plastics.



Trimesic Acid

Trimesic acid is also referred to as 5-carboxyisophthalic acid [554-95-0] trimesinic acid, or trimesitinic acid. It is a small-volume, synthetic chemical and is sold commercially. Traces of trimesic acid as well as other aromatic carboxylic acids with three or more carboxylic acid groups are found in lignite (137), and when various types of coals or coal components such as brown coal, asphaltene, or coal-tar pitch are oxidized.

Manufacture. The only current U.S. manufacturer of trimesic acid is Amoco Chemical Co. It is produced by oxidation of mesitylene (1,3,5-trimethylbenzene) via the liquid-phase oxidation in acetic acid using the cobalt– manganese–bromine catalyst system (138). This is a variant of the system used to produce terephthalic and isophthalic acids as well as trimellitic anhydride. American Bio-Synthetics Corp. did produce it by batch oxidation of mesitylene with potassium permanganate.

Economic Aspects. Trimesic acid being a small-volume chemical, availability is in polyethylene-lined fiber drums in 34- and 136-kg quantities. Mesitylene [108-67-8], the raw material, is readily available, and oxidation to trimesic acid presents no problems. The market size is dictated by commercial uses.

Specifications and Standards. Table 35 lists specifications for trimesic acid as produced by Amoco Chemical (139). Typically, the product is over 98% pure.

Table 35. Specifications for Trimesic Acid

Property	Specification	Test method
physical appearance	free flowing powder	
freezing point, °C	374–376	cryoscopy
acid number, mg KOH/g	786 min	titration
organic impurities, wt %	0.5 max	chromatography
volatiles, wt %	0.5 max	ASTM E203

ash, wt %	0.15 max	combustion
-----------	----------	------------

Health and Safety Factors. Trimesic acid is an irritant to the skin, eyes, and respiratory system (140). It is mildly toxic when ingested. The oral LD₅₀ in rats has been reported as 8.4 g/kg (141). Trimesic acid is flammable, and precautions similar to those noted for terephthalic acid and isophthalic acid as regards dust clouds and fire extinguishing agents should be followed.

Uses. The only significant commercial use for trimesic acid is as a cross-linking agent in solid rocket fuels. Other reported uses include cross-linking agents for polymers (142), as the base acid in making triesters as plasticizers (143), as a component of electrostatic toners (144), and as a stationary phase for gas chromatography (145,146). Trimesic acid is sold as a research chemical (99).

Hemimellitic Acid

Hemimellitic acid is not manufactured commercially but is available as a laboratory chemical in a hydrate form (99). Like trimesic acid, it is formed when coal-like materials are oxidized, but can be synthesized in a purer form by oxidizing hemimellitene [562-73-8] (1,2,3-trimethylbenzene) or hemimellitol [526-85-2]. Hemimellitic anhydride can be produced by thermal dehydration of the acid in trichlorobenzene at 261°C (147). Synthesis of hemimellitic acid and anhydride have been described (148). There are no reported uses which are unique to hemimellitic acid.

Pyromellitic Acid and Pyromellitic Dianhydride

Pyromellitic acid is a commercial product, and it forms a dianhydride which has specialized commercial applications, primarily as an ingredient in the preparation of high temperature polymers. The IUPAC name of the dianhydride is 1*H*,3*H*-benzo(1,2-*c*4,5-*c'*)difuran-1,3,5,7-tetrone.

Physical and Chemical Properties. Tables 36 and 37 contain some of the physical and chemical properties of pyromellitic acid and its anhydride.

Table 36. Physical Contants of Pyromellitic Acid and Pyromellitic Dianhydride

Property	Pyromellitic acid	Pyromellitic dianhydride
mp, °C	281–284.5	284–286
bp, °C		380–400
vapor pressure, kPa ^a		
at 290°C		8.4
350°C		44
heat of vaporization, kJ/mol ^b		83.4
specific gravity at 20°C		1.680
heat of combustion at 25°C, kJ/mol ^b		3313
heat of formation at 25°C, kJ/mol ^b	1571	907
heat of fusion at 285°C, kJ/mol ^b		15.8

^a To convert kPa to mm Hg, multiply by 7.5.

^b To convert J to cal, divide by 4.184.

Table 37. Solubilities of Pyromellitic Acid and Pyromellitic Dianhydride^a

Solvent	Temperature, °C		
	25°C	50°C	70°C
pyromellitic acid			
water ^b	1.5	5.0	12
ethanol	15	22	30
dimethylformamide		31	
dimethyl sulfoxide ^c			
pyromellitic dianhyhdride			
acetone	7.5	8.5	
dimethylformamide	18	35	72
dimethyl sulfoxide	24	43	80

^a g/100 g solvent.

^b 30 g/100 g H₂O at 90°C.

^c 62 g/100 g DMSO at 100°C.

Manufacture and Processing. Pyromellitic acid and its dianhydride can be synthesized by oxidizing durene [95-93-2] (1,2,4,5-tetramethylbenzene). Liquid-phase oxidation using strong oxidants such as nitric acid, chromic acid, or potassium permanganate produces the acid which can be dehydrated to the dianhydride in a separate step. This technology is practiced by Allco Chemical Co., a part of International Specialty Chemicals.

The use of the liquid-phase process in acetic acid with the cobalt–manganese–bromine system as explained in the terephthalic acid section is also possible (149). This process has been used by Amoco Chemical to produce pyromellitic acid, and facilities remain in place to do so again in the future. As with all liquid-phase oxidations of this type, yields are high. A separate dehydration step would be needed to yield the dianhydride.

Mitsubishi Gas Chemical Co. in Japan produces pyromellitic dianhydride in the same unit used for trimellitic anhydride production (105). This process starts with pseudocumene, which is first carbonylated with carbon monoxide in the presence of boron trifluoride and hydrogen fluoride to form 2,4,5-trimethylbenzaldehyde. The liquid-phase oxidation of the trimethylbenzaldehyde to pyromellitic acid and subsequent processing steps are much the same as described for the Mitsubishi Gas Chemical process in the trimellitic acid section. The production of pyromellitic anhydride is in conjunction with a joint venture agreement with Du Pont.

Vapor-phase catalytic oxidation of durene is a more direct route to the dianhydride. Huls in Europe apparently uses this route, which eliminates the need for a separate dehydration step and for handling of any oxidants or solvents. Continuous operation is facilitated, corrosion is minimized, and product recovery is simplified. The vapor-phase oxidation of durene is similar to that of *o*-xylene to phthalic anhydride, and phthalic anhydride units can be

converted to produce pyromellitic dianhydride. Production of the durenene with two anhydride rings and a lower vapor pressure than phthalic anhydride is more difficult, however. In theory, the vapor-phase oxidation does not result in the formation of carboxyl groups that are subsequently dehydrated by the same catalyst. It proceeds through an intramolecular, five-membered ring which oxidizes to 4,5-dimethylphthalic anhydride. This is the presumed intermediate in the production of pyromellitic dianhydride (150). Catalysts specifically noted for durenene oxidation contain vanadium, titanium, tin, zirconium, phosphorus, niobium, potassium, cesium, rubidium, tellurium, and antimony (151). Honeycomb supports with titanium and vanadium catalysts are also noted (152).

Production, Storage, and Shipment. As noted above, Alco Chemical, Amoco Chemical, Mitsubishi Gas Chemical, and Huls all produce either the acid or the anhydride using different production techniques. The relatively small production volumes of pyromellitic acid and dianhydride results in both storage and shipment in polyethylene-lined fiber drums of 22–136-kg capacity.

Economic Aspects. Prices for pyromellitic acid were about \$14 /kg in 1994. The dianhydride sold for about \$19–25 /kg depending on purity, and prices of the dianhydride ground to a fine 3-μm size were \$2 /kg higher (153). Production amounts are not released and are dictated by market needs. The use of some multipurpose units to make this product means that the amounts produced are highly variable.

Specifications and Standards. Typical properties of pyromellitic acid are given in Table 38 (154,155). In many cases, specifications are negotiated between manufacturer and customer based on the needs of the final product.

Table 38. Specifications for Pyromellitic Dianhydride

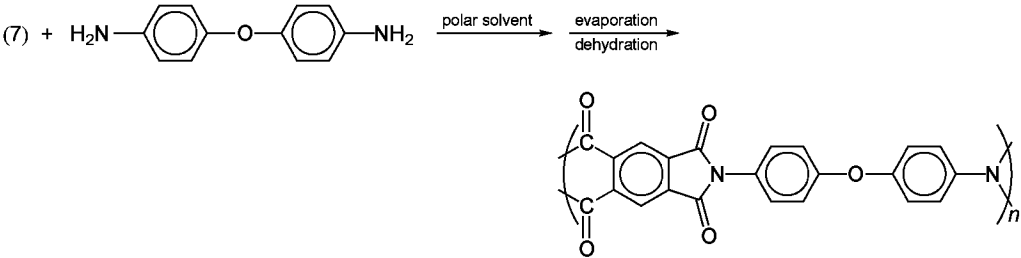
Property	Specification	Test method
purity, % min	99.5	by difference
organic impurities, wt %	1.5 max	chromatography
color	almost white	visual
mp, °C	282 min	cryoscopy
ash, ppm	400 max	combustion
volatiles, wt %	0.5 max	evaporation
bp ^a , °C	397–400	
bp ^b , °C	305–310	

^a At 101.3 kPa (1 atm).

^b At 4 kPa (0.04 atm).

Health and Safety Factors. Both pyromellitic acid and its dianhydride irritate skin, eyes, and mucous membranes, and they cause skin sensitization (156). When it comes in contact with moist tissue the dianhydride converts to the acid. Direct contact with should be avoided and protective clothing should be worn in areas where it is used. The LD₅₀ for intergastric administration in rats is 2.2–2.6 g /kg (157). In 6-mo experiments, the maximum nontoxic dose was 0.07 mg /kg d, and it affected the liver, kidney, and reproductive tract. Precautions against fire and dust explosions as explained in the terephthalic acid section should be followed.

Uses. Pyromellitic dianhydride imparts heat stability in applications where it is used. Its relatively high price limits its use to these applications. The principal commercial use is as a raw material for polyimide resins (see POLYIMIDES). These polypyromellitimides are condensation polymers of the dianhydride and aromatic diamines such as 4,4'-oxydianiline:



Du Pont produces this polymer under the trade names of Kapton, Pyralin, Vespel, and Pyre-ML. The trade names refer to polyimides used for film, semiconductor coatings, molding applications, and wire enamel, respectively. They have excellent thermal, electrical, and physical properties.

Because the heat distortion temperature of cured epoxy resins (qv) increases with the functionality of the curing agents, pyromellitic dianhydride is used to cross-link epoxy resins for elevated temperature service. The dianhydride may be added as a dispersion of micropulverized powder in liquid epoxy resin or as a glycol adduct (158). Such epoxies may be used as an insulating layer in printed circuit boards to improve heat resistance (159). Other uses include inhibition of corrosion (160,161), hot melt traffic paints (162), azo pigments (163), adhesives (164), and photoresist compounds (165).

Mellophanic and Prehnitic Acids

Neither mellophanic acid nor prehnitic acid are commercial products, nor are they available as laboratory chemicals. Some references identify the 1,2,3,4-isomer as prehnitic acid, and the 1,2,3,5-isomer as mellophanic acid; this designation relates to the name prehnitene [488-23-3], which is used for 1,2,3,4-tetra-methylbenzene. However, *Chemical Abstracts* identifies the 1,2,3,4-isomer as mellophanic acid and the 1,2,3,5-isomer as prehnitic acid. Both acids could be synthesized by oxidizing the corresponding tetramethylbenzenes. Mellophanic acid has been made by the ring dehydrogenation of 1,2,3,4-cyclohexanetetracarboxylic acid with bromine (165). Prehnitic acid has been synthesized by the cobalt catalyzed carbonylation of Schiff bases (166). These acids are also formed in the oxidation of coal-like materials.

Mellophanic acid and prehnitic acid have few literature references for uses. Possible applications are as a tanning agent in combination with an aluminum salt (167), to give a pearl-like gloss to polyester–polymethacrylate resins (168) or brightness to polypropylene fibers (169), or to increase detergent efficiency (170). Mellophanic dianhydride has a relatively low melting point and good solubility in epoxy resins. It has therefore been considered as a substitute for pyromellitic dianhydride as a cross-linking agent (171).

Benzenepentacarboxylic Acid and Mellitic Acid

Neither benzenepentacarboxylic acid nor mellitic acid are manufactured commercially, but synthetic mellitic acid can be purchased as a laboratory chemical (99). Both can be synthesized by oxidizing the corresponding methylbenzenes or other substituted benzenes, and both are present in trace amounts after oxidation of coal or coal-like substances.

Both the acids, their sodium salts, as well as benzenetricarboxylic and benzenetetracarboxylic acids have been considered as substitutes for phosphorus- and nitrogen-containing detergent builders (172). Although satisfactorily biodegradable, these acids are not as effective as other builders and have not been used (173). Other potential uses for mellitic acid are as part of the resin component of sand molds (174), as an additive to anodizing baths for improving finishes on aluminum (175), and as an additive to oral antibacterial preparations for preventing tooth discoloration (176). Both

benzenepentacarboxylic acid and mellitic acid have been mentioned as ingredients in bonding agents (177) and dental fillings (178).

BIBLIOGRAPHY

"Phthalic Acids" in *ECT* 1st ed., Vol. 10, pp. 584–607, by R. I. Sturton; "Phthalic Acids and Other Benzenepolycarboxylic Acids" in *ECT* 2nd ed., Vol. 15, pp. 444–487, by P. H. Towle, R. H. Baldwin, and D. H. Meyer, Amoco Chemicals Corp.; in *ECT* 3rd ed., Vol. 17, pp. 732–777, by A. G. Bemis, J. A. Dindorf, B. Horwood, and C. Samans, Amoco Chemicals Corp.

1. *Beilsteins Handbuch Der Organischen Chemie*, 4, Aufl., Bd. IX 5, Springer-Verlag, Berlin, Germany, 1971.
2. R. C. Weast and J. G. Grasselli, eds., *Handbook of Data on Organic Compounds*, 2nd ed., CRC Press, Boca Raton, Fla., 1989.
3. *Dictionary of Organic Compounds*, 4th ed., Oxford University Press, New York, 1965.
4. I. Puskas and E. K. Fields, *Ind. Eng. Chem. Prod. Res. Dev.* **9**, 403 (1970).
5. **U.S. Pat. 1,285,117** (Nov. 19, 1918), H. D. Gibles and C. Conover.
6. **U.S. Pats. 1,787,416** and **1,787,417** (Dec. 20, 1930), A. Wohl (to I. G. Farbenindustrie).
7. **U.S. Pat. 1,971,188** (Aug. 28, 1934), A. Wohl (to I. G. Farbenindustrie).
8. *Oil Paint Drug Rep.* **189**, 3 (Mar. 14, 1966).
9. I. E. Levine, *The Chemistry of Petroleum Hydrocarbons*, Vol. 3, Reinhold Publishing Corp., New York, 1955, Chapt. 38.
10. W. G. Toland and E. L. Nemes, *Oil Gas* **54**, 232 (July 25, 1955).
11. **U.S. Pat. 2,245,528** (June 10, 1941), D. J. Loder (to Du Pont).
12. **U.S. Pat. 2,425,398** (Aug. 10, 1947), R. P. Ruthruff (to Sherwin Williams).
13. V. Nikolov, A. Anastanov, and K. Kussurski, *Chim. Ind. (Milan)* **73**(2), 111 (1991).
14. O. Wiedemann and W. Gierer, *Chem Eng.* **86**(3), 62 (1979).
15. R. F. Schwab and W. H. Doyle, *Chem. Eng. Prog.* **66**(9), 49 (1970).
16. M. S. Wainwright and N. R. Foster, *Catal. Rev. Sci. Eng.* **19**, 211 (1979).
17. **U.S. Pat. 4,405,505** (Sept. 20, 1983), A. Neri, L. Capitanio, and G. Stefani (to Alusuisse Italia).
18. **U.S. Pat. 4,036,783** (July 19, 1977), K. Blechschmitt and co-workers (to BASF).
19. **Ger. Pat. 2,546,268** (Apr. 21, 1977), K. Blechschmitt and co-workers (to BASF).
20. **Fr. Pat. 2,332,276** (June 17, 1977), (to Nippon Shokubai Kagaku Kogyo).
21. T. G. Smith and J. J. Carberry, *J. Chem. Eng.* **53**, 347 (1975).
22. P. Sadlukan and E. E. Petersen, *AIChE J.* **22**, 808 (1976).
23. **Jpn. Kokai 79 19,479** (Feb. 14, 1979), M. Takada, H. Uhara, and T. Sako (to Nippon Shokubai Kagaku Kogyo).
24. J. J. Graham and co-workers, *Chem. Eng. Prog.* **58**(1), 96 (1962).
25. C. O. Bolthrunis, *Chem. Eng. Prog.* **85**(5), 51 (1989).
26. J. J. Graham, *Chem. Eng. Prog.* **66**(9), 54 (1970).
27. *Hydrocarbon Process.* **56**(11), 196 (1977).
28. **U.S. Pat. 4,062,871** (Dec. 13, 1977), G. Gehrken and G. Kenrecke (to Chemibau).
29. *Hydrocarbon Process.* **58**(11), 190 (1979).
30. **U.S. Pat. 4,165,324** (Aug. 21, 1979), H. Schroeder and S. J. Kulpa (to Standard Oil Indiana).
31. V. Nikolov and A. Anastasov, *AIChE J.* **35**, 511 (1989).
32. *Chem. Mkt. Rep.* **244**, 13 (Oct. 18, 1993).
33. *Chem. Mkt. Rep.* **242**, 42 (Aug. 10, 1992).
34. *Chemical Safety Data Sheet SD-61 for Phthalic Anhydride*, Manufacturing Chemists Association, Washington, D.C., 1956.
35. M. Sittig, *Hazardous and Toxic Effects of Industrial Chemicals*, Noyes Data, Park Ridge, N.J., 1979.
36. G. Moscato, G. Fherson, A. Salvaterra, and I. Vidi, *G. Ital. Med. Lav.* **8**(2), 57 (1987).
37. I. W. Fawcett, A. J. Taylor, A. J. Newman, and J. Pepys, *Clin. Allergy* **7**(1), 1 (1977).
38. J. A. Sagan, *Biul. Sluzby Sanit. Epidemiol. Wojewodztwa Katowickiego* **17**, 359 (1973).
39. N. I. Sax and R. J. Lewis, eds., *Dangerous Properties of Industrial Materials*, 7th ed., Van Nostrand Reinhold, New York, 1989, p. 2783.
40. J. Nagy, H. G. Dorsett, and M. Jacobson, *U.S. Bur. Mines Rep. Invest.*, 6543 (1964).
41. M. Jacobson, A. R. Cooper, and J. Nagy, *U.S. Bur. Mines Rep. Invest.*, 5971 (1962).
42. *Chem. Mkt. Rep.* **242**, 42 (July 27, 1992).
43. **Brit. Pat. 655,074** (July 11, 1951), L. A. Burrows, R. M. Cavanaugh, and W. M. Nague (to Du Pont).
44. **U.S. Pat. 2,636,899** (Apr. 28, 1953), L. A. Burrow, R. M. Cavanaugh and J. E. Lufkin (to Du Pont).
45. **Brit. Pat. 895,926** (May 9, 1962) (to Du Pont).
46. **Brit. Pat. 712,315** (July 21, 1954), C. Gardner, G. M. Henderson, and E. C. Owen (to Imperial Chemical Industries).
47. **Brit. Pat. 758,676** (Oct. 10, 1956), C. Gardner and E. C. Owen (to Imperial Chemical Industries).
48. **U.S. Pat. 2,459,014** (Jan. 11, 1949), R. M. Cavanaugh and J. E. Lufkin (to Du Pont).
49. H. Lueckert, *Fiber Produce.* 8–14 (Apr. 1981).
50. A. P. Aneja and V. P. Aneja, *Chem. Econ. Eng. Rev.* **10**(9), 33 (1978).
51. **U.S. Pat. 2,833,816** (May 6, 1958), R. S. Barker and A. Saffer (to Mid-Century Corp.)
52. **U.S. Pat. 3,089,906** (May 14, 1963), R. S. Barker and A. Saffer (to Mid-Century Corp.)
53. T. G. Otchy and K. J. Herbert, "First Large Scale Catalytic Oxidation System for PTA Plant CO and VOC Abatement," *85th Annual Air and Waste Management Association Meeting*, Air & Waste Management Association, Pittsburgh, Pa., 1992.
54. S. Shelley, *Chem. Eng.* **98**(12), 90 (1991).
55. *Hydrocarbon Process.* **56**(11), 149 (1977).
56. *Eastman News*, **43**(14) (1988).
57. *Hydrocarbon Process.* **56**(11), 230 (1977).
58. H. S. Bryant, C. A. Duval, L. E. McMakin, and J. I. Savoca, *Chem. Eng. Prog.* **67**(9), 69 (1971).
59. **Fr. Pat. 1,560,461** (Mar. 21, 1969) (to Toray).
60. **U.S. Pat. 2,905,709** (Sept. 22, 1959), W. Schenk and G. Schiller (to Henkel & Cie.).
61. **U.S. Pat. 2,823,229** (Feb. 11, 1958), B. Raecke (to Henkel & Cie.).
62. **Brit. Pat. 975,113** (Nov. 11, 1964) (to Teikoku).
63. **Rus. Pat. 69 20,525** (Oct. 13, 1967) (to Teijin).
64. **Rus. Pat. 69 20,091** (Aug. 29, 1969) (to Teijin).
65. **Rus. Pat. 72 24,549** (July 6, 1972), K. Okuno and co-workers (to Mitsubishi Chemical).
66. **Rus. Pat. 72 19,537** (June 3, 1972), K. Okuno and Y. Isoga (to Mitsubishi Chemical Research).
67. **U.S. Pat. 2,653,165** (Sept. 22, 1953), I. E. Levine (to California Research).
68. **U.S. Pat. 2,772,305** (Nov. 27, 1956), I. E. Levine and W. G. Toland (to California Research).
69. **Ger. Pat. 1,070,615** (May 25, 1960), E. Katzschmann (to Chemische Werke Witten).

70.

H. J. Korte, H. Schroeder, and A. Schoengen, "The PTA Process of Huls Troisdorf AG," *1988 AIChE Summer National Meeting*, American Institute of Chemical Engineers, New York.

71.

U.S. Pat. 3,253,017 (May 24, 1966), E. Katschmann (to Chemische Werke Witten).

72.

Hydrocarbon Process. **58**(11), 154 (1979).

73.

Hydrocarbon Process. (Nov. 1977).

74.

Fr. Pat. 2,163,472 (July 27, 1973), G. Hoffmann, R. Cordes, and W. Merkel (to Dynamit Nobel).

75.

U.S. Pat. 2,976,030 (Mar. 21, 1961), D. H. Meyer (to Standard Oil Indiana); **Brit. Pat. 1,202,560** (Aug. 19, 1970) (to Toray).

76.

U.S. Pat. 2,491,660 (Dec. 20, 1949), W. F. Gresham (to Du Pont).

77.

U.S. Pat. 3,584,039 (June 8, 1971), D. H. Meyer (to Standard Oil Indiana).

78.

D. H. Meyer, H. P. Pohlmann, and C. E. Johnson, "Production of Aromatic Acids by the Amoco Chemicals Corporation," *Pacific Chemical Engineering Congress*, Kyoto, Japan., Oct. 1972.

79.

U.S. Pat. 3,931,305 (Jan. 6, 1976), J. A. Fisher (to Standard Oil Indiana).

80.

H. Hizikata, *Chem. Econ. Eng. Rev.* **9**(9), 32 (1977).

81.

K. Matsuzawa, *Chem. Econ. Eng. Rev.* **8**(9), 25 (1976).

82.

U.S. Pat. 4,877,900 (Oct. 31, 1989), T. Tamaru and Y. Izumisawa (to Mitsubishi Chemical Industries Limited).

83.

U.S. Pat. 4,447,646 (May 8, 1984), G. I. Johnson and J. E. Kiefer (to Eastman Kodak Co.).

84.

D. Savostianoff and E. R. Didier, *Informations Chem.* **352**, 121–129 (Oct. 1993).

85.

Chem. Mkt. Rep. **236**, 46 (July 17, 1989).

86.

J. V. Marhold, *Sb. Vys. Toxicol. Vys. Latid A Prinavku*, 52 (1972).

87.

A. E. Moffitt, J. J. Clary, T. R. Lewis, M. D. Blank, and V. B. Perone, *Amer. Ind. Hyg. Assoc. J.* **36**, 633 (1975).

88.

H. d'A. Heck, *Banbury Rep.* **25**, 233 (1987).

89.

Material Safety Data Sheet, Amoco TA-33, Amoco Chemical Co., Chicago, Ill., Dec. 1993.

90.

W. J. Krasavage, *Amer. Ind. Hyg. Assoc. J.* **34**, 455 (1973).

91.

A. D. Craven and M. G. Foster, *Combust. Flame* **11**, 408 (1967).

92.

Chem. Week **153**, 27 (Dec. 22–29, 1993).

93.

L. A. Carmichael, *Aromatic Amines*, Rpt. No. 76 (Private Subscription), Stanford Research Institute, Menlo Park, Calif. 1972, pp. 159–194.

94.

Eur. Chem., 538 (1978).

95.

Belg. Pat. 871,044 (Feb. 1, 1979), (to Societa Italiana Serie Acetica Sintetica).

96.

Ref. 39, p. 2401.

97.

Bulletin IP-43b, Amoco Chemical Co., Chicago, Ill., 1989.

98.

Rus. Pat. 78 40,048 (Apr. 12, 1978), H. Matsunaga and co-workers (to Sumitomo Chemical).

99.

Catalog Handbook of Fine Chemicals, Aldrich Chemical Co., Milwaukee, Wis., 1994.

100.

WO Pat. 91-08,204 (June 13, 1991), C. M. Park, R. Coates, J. K. Holzhauer, and J. V. Peterson (to Amoco).

101.

U.S. Pat. 5,004,830 (Apr. 2, 1991), C. M. Park and W. P. Schammel (to Amoco).

102.

U.S. Pat. 4,992,579 (Feb. 12, 1991), W. P. Schammel (to Amoco).

103.

U.S. Pat. 4,895,978 (Jan. 23, 1990), J. K. Darin and A. G. Bemis (to Amoco).

104.

M. Komatsu, *Nippon Kagaku Kyokai Geppo*, 11–17 (Apr. 1988).

105.

T. Kasahara and Y. Takamizawa, *Kagaku Kogaku*, **52**(6), 416 (1988).

106.

L. M. Elkin, *PEP Review 85-1-1*, SRI International (private subscription), Menlo Park, Calif., Dec. 1985.

107.

Rus. Pat. 83 121,244 (July 19, 1983), T. Tanaka and co-workers (to Mitsubishi Gas Chemical).

108.

U.S. Pat. 4,500,730 (Feb. 19, 1985), T. Tanaka, M. Hataya, K. Tanaka, Y. Sakai, and Y. Hamada (to Mitsubishi Gas Chemical).

109.

Chem. Mkt. Rep. **240**, 4 (Dec. 30, 1991).

110.

Eur. Pat. 0 513 835 A2 (Nov. 19, 1992), C. Fumagalli, L. Capitanio, and G. Stefani (to Alusuisse Italia).

111.

U.S. Pat. 4,398,040 (Aug. 9, 1983), T. Suzuki, K. Kitahara, S. Naito, and T. Tsuji (to Mitsubishi Gas Chemical).

112.

U.S. Pat. 5,171,881 (Dec. 15, 1992), S. H. Park, J. S. Go, and J. W. Sim (to Yukong, Korea).

113.

Rus. Pat. 89 121,238 (May 12, 1989), M. Kodama, T. Yoshii, and K. Kawamura (to Idemitsu Petrochemical).

114.

Bulletin TM-143, Amoco Chemical Co., Chicago, Ill., Aug. 1993.

115.

"Plasticizers", *Chemical Economics Handbook*, 1992; p. 579.5000 technical data, SRI International, Menlo Park, Calif., 1994.

116.

Bulletin TM-144, Amoco Chemical Co., Chicago, Ill., Aug. 1993.

117.

Material Safety Data Sheet, Amoco TMA, Amoco Chemical Co., Chicago, Ill., June 1989.

118.

Bulletin TM-135, Amoco Chemical Co., Chicago, Ill., Feb. 1991.

119.

Bulletin GTSR 58a-TM, Amoco Chemical Co., Chicago, Ill., Mar. 1989.

120.

Bulletin GTSR-98, Amoco Chemical Co., Chicago, Ill., Aug. 1992.

121.

Bulletin IP-66, Amoco Chemical Co., Chicago, Ill., July 1978.

122.

Eur. Pat. 384,578 (Aug. 29, 1990), J. B. Simons and B. L. Hedrick (to Morton International).

123.

Bulletin TMA-128a, Amoco Chemical Co., Chicago, Ill., Oct. 1981.

124.

Rus. Pat. 90 167,561 (June 27, 1990), H. Kobayashi and T. Kamitaki (to Canon KK).

125.

Eur. Pat. 0 409,557 (Jan. 23, 1991), S. Kubo and co-workers (to Mitsubishi Rayon).

126.

Alcolac Technical Information, SIPOMER TATM, Linthicum, Md.

127.

Bulletin IP-65c, Amoco Chemical Co., Chicago, Ill., Oct. 1992.

128.

Bulletin GTSR-105, Amoco Chemical Co., Chicago, Ill., Aug. 1993.

129.

Bulletin TM-108c, Amoco Chemical Co., Chicago, Ill., Oct. 1989.

130.

Bulletin TM-124c, Amoco Chemical Co., Chicago, Ill., Oct. 1989.

131.

Bulletin TM-139, Amoco Chemical Co., Chicago, Ill., Oct. 1990.

132.

U.S. Pat. 4,983,716 (Jan. 8, 1991), M. Rao, R. F. Tomko, and D. R. Sayre (to Sherwin Williams).

133.

Ger. Pat. 3,817,614 (Aug. 17, 1989), D. Winkler, W. Schuelke, G. Hoeveler (to Herberts GmbH).

134.

Rus. Pat. 91 07,775 (Jan. 14, 1991), S. Shigeharu (to Furukawa Electric).

135.

U.S. Pat. 4,496,715 (Jan. 25, 1985), F. A. Sattler (to General Electric).

136.

Bulletin TMA-3b, Amoco Chemical Co., Chicago, Ill., June 1983.

137.

R. Hayatsu and co-workers, *Nature* **275**, 116 (1978).

138.

U.S. Pat. 5,030,749 (July 9, 1991), G. P. Hussmann (to Amoco Corp.).

139.

"Amoco Trimesic Acid," *Bulletin FA-1*, Amoco Chemical Co., Chicago, Ill.

140.

Trimesic Acid Material Safety Data Sheet, Amoco Chemical Co., Chicago, Ill., Feb. 27, 1989.

141.

Trimesic Acid Product and Safety Data Sheets, American Bio-Synthetics Corp., Milwaukee, Wis.

142.

U.S. Pat. 4,602,050 (July 22, 1986), R. C. West and P. I. Djurovich (to Wisconsin Alumni Research Foundation).

143.

Eur. Pat. Appl. 49,047 (Apr. 7, 1982), T. Le Marre Patton (to Exxon Research and Engineering).

144.

Rus. Pat. 86 07,846 (Jan. 14, 1986), Y. Kamijama and K. Ito (to Hitachi Metals, Ltd.).

145.

F. Mangani, A. Fabbri, and G. Crescentini, *Anal. Chem.* **58**, 3261 (1986).

146.

G. Cosmi, A. Di Corcia, R. Samperi, and G. Vinci, *Chromatographia* **16**, 322 (1982).

147.

Rus. Pat. 69 02,454 (Feb. 1, 1969), H. Miyamori (to Mitsubishi Edogawa Chemical).

148.

L. D. Kalinina, B. V. Suvorov, and K. V. Prokof'ev, *Tr. Inst. Khim. Nauk. Akad. nauk. Kaz. SSR* **45**, 108 (1977).

149.

W. H. Scheeline, *Trimellitic and Pyromellitic Anhydrides*, Report No. 66 (private subscription), Stanford Research Institute, Menlo Park, Calif., 1970.

150.

U.S. Pat. 4,665,200 (May 12, 1987), Y. Nakanishi and co-workers (to Nippon Shokubai Kagaku Kogto Co.).

151.

Rus. Pat. 88 242,349 (Oct. 7, 1988), N. Enomoto, Y. Naruse, and Y. Asami (to Kawasaki Steel).

152.

Rus. Pat. 86 27,942 (Feb. 7, 1986), M. Suematsu, S. Otomo, and K. Nakaoka (to Toray Industries).

153.

Technical data, Allco Chemical Co., Galena, Kans., 1994.

154.

Amoco PMA, *Bulletin FA-8*, Amoco Chemical Co., Chicago, Ill.

155.

Pyromellitic Dianhydride, Safety Data Sheet and Brochure, Veba-Chemie AG, Gelsenkirchenbuer, Germany, 1978.

156.

Pyromellitic Anhydride Material Safety Data Sheet, Amoco Chemical Co., Chicago, Ill., June 18, 1987.

157.

V. A. Kondratyuk and co-workers, *Gig. Sanit.* (11), 79 (1986).

158.

W. Brenner, D. Lum, and M. W. Riley, *High Temperature Plastics*, Reinhold Publishing Corp., New York, 1962, pp. 29–37.

159.

Rus. Pat. 91 30,396 (Feb. 8, 1991), S. Demura, R. Otake, and M. Hayashi (to Dainippon Ink and Chemicals).

160.

Rus. Pat. 76 93,741 (Aug. 17, 1976), T. Suzuki and A. Maeda (to Kurita Water Industries).

161.

S. M. Hassan, M. N. Moussa, M. M. El-Tagoury, and A. A. Radi, *Anti-Corros. Methods Mater.* **37**(2), 8 (1990).

162.

Rus. Pat. 77 52,934 (Oct. 25, 1977), K. Toshima, A. Ishikawa, and H. Inoue (to Nippon Zeon).

163.

Czech. Pat. 146,005 (Nov. 15, 1972), L. Havlickova and S. Kmonickova.

164.

Rus. Pat. 89 255,854 (Oct. 12, 1989) T. Asos (to Fuji Photo Film).

165.

Rus. Pat. 70 36,901 (Nov. 24, 1970), S. Gomi, M. Takahashi, and H. Nagai (to Kureha Chemical Industry).

166.

T. Kajimoto and J. Tauji, *J. Org. Chem.* **48**, 1685 (1983).

167.

Brit. Pat. Appl. GB 2,153,844 (Aug. 29, 1985), S. Dasgupta (to Yorkshire Chemical PLC).

168.

Rus. Pat. 72 17,916 (May 24, 1972), S. Torikai, T. Muraki, and Y. Shin (to Toray).

169.

Rus. Pat. 71 12,906 (Apr. 2, 1971), M. Yamata and M. Nakamura (to Dainippon Seika Kogyo).

170.

Neth. Pat. App. 74 07,851 (Dec. 17, 1974) (to Farbwerke Hoechst).

171.

U.S. Pat. 3,444,135 (May 13, 1969), N. P. Greco and C. J. Pennino (to Koppers).

172.

Can. Pat. 950,788 (July 8, 1974), D. S. Connor and J. E. Thompson (to Proctor and Gamble).

173.

F. J. Letkiewicz, NTIS PB-248, U.S. Environmental Protection Agency, Washington, D.C., 1975, p. 835.

174.

Rus. Pat. 90 295,633 (Dec. 6, 1990), M. Suzuki and co-workers (to Gunei Chemical Industry Co., Ltd.)

175.

Ger. Pat. 2,650,989 (May 11, 1978), H. J. Gowhausen, W. Kirchoff, and T. Lindener (to Henkel).

176.

Ger. Pat. 2,755,847 (June 29, 1978) A. Gaffar (to Colgate Palmolive).

177.

Rus. Pat. 82 164,175 (Oct. 8, 1982) (to Japan Synthetic Rubber Co. Ltd.).

178.

Rus. Pat. 82 02,209 (Jan. 7, 1982) (to G-C Dental Industrial Corp.).

Chang-Man Park

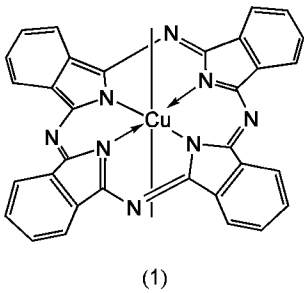
Richard J. Sheehan

Amoco Chemical Company

PHTHALOCYANINE COMPOUNDS

Phthalocyanine[574-93-6], C₃₂H₁₈N₈, compounds have found widespread acceptance in a variety of applications. The discovery of iron phthalocyanine [132-16-1] and the elucidation of its structure led to the commercial application of copper phthalocyanine [147-14-8] (1).

Copper phthalocyanine (1) was developed in the 1930s and is the most commonly used blue organic pigment in the coatings (qv), paint (qv), and printing inks (qv) industry. Phthalocyanine forms complexes with numerous metals. Various complexes with 66 chemical elements are known (1–7). The coordination number of the metal in the phthalocyanine complexes ranges from 4 (Cu, Li, Pd) to 5 (U), or 6 (metals with two additional ligands such as H₂O or NH₃) to 8 (Nd(III) HPc₂ or U(IV) Pc₂) (8–13), where Pc represents the phthalocyanine ligand. Polymeric phthalocyanines can also be synthesized (Si, Gc, Sn) (14–17). Metal-free phthalocyanine, H₂Pc, in which the central copper atom in (1) is replaced by two hydrogen atoms, has interesting photosensitive and semiconductor properties (1). Phthalocyanines are structurally related to naturally occurring dyes such as hemoglobin and chlorophyll A (18).



Physical Properties

The density of β-phthalocyanine, H₂Pc, is 1.43 g cm⁻³; β-copper phthalocyanine [14832-14-5], CuPc, 1.61 g cm⁻³; and polychloro-copper phthalocyanine, 2.14 g cm⁻³. The color of most phthalocyanines ranges from blue-black to a metallic bronze, depending on the manufacturing process and the chemical and crystalline form of the material (19). The colors of the finely divided pigment forms vary from dark blue to green, as phthalocyanines absorb in the visible region at 600–700 μm (2). Most compounds do not melt but sublime above 200°C. CuPc can be sublimed without decomposition at 500–580°C under an inert gas and normal pressure and at 900°C under vacuum (20). It decomposes vigorously, however, at 405–420°C in air and in nitrogen between 460–630°C (21,22). The thermodynamic stability of the five crystalline forms of CuPc increases in the sequence α γ < δ < ε < β(23,24). The solubility of most phthalocyanines in water and organic solvents is very low. The α-form, however, is slightly soluble in polar solvents and converts rapidly to the β-form (25,26). Three modifications of H₂Pc, α, β (27,28), and X (29) have been found. Polychloro CuPc has only an α-modification (30). The arrangement of the planer stackings differs as illustrated by the structures of α- and β-CuPc in Figure 1 (27,28).

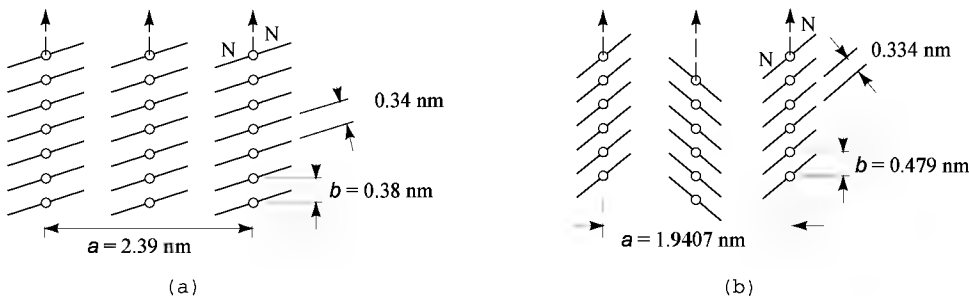
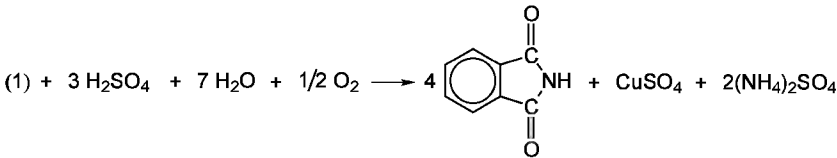


Fig. 1. Arrangement of the copper phthalocyanine molecule in the (a) α- and (b) β-forms.

Chemical Properties

The chemical properties of phthalocyanines depend mostly on the nature of the central atom. Phthalocyanines are stable to atmospheric oxygen up to approximately 100°C. Mild oxidation may lead to the formation of oxidation intermediates that can be reduced to the original products (29). In aqueous solutions of strong oxidants, the phthalocyanine ring is completely destroyed and oxidized to phthalimide. Oxidation in the presence of ceric sulfate can be used to determine the amount of copper phthalocyanine quantitatively (30).



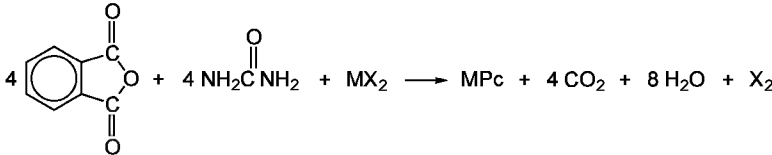
Oxidation can also occur at the central metal atom of the phthalocyanine system (2). Mn phthalocyanine, for example, can be produced in these different oxidation states, depending on the solvent (2,31,32). The carbon atom of the ring system and the central metal atom can be reduced (33), some reversibly, eg. in vatting (34–41). Phthalocyanine compounds exhibit favorable catalytic properties which makes them interesting for applications in dehydrogenation, oxidation, electrocatalysis, gas-phase reactions, and fuel cells (qv) (1,2,42–49).

Manufacturing and Processing

Phthalocyanine compounds have been synthesized with various metals (1,2,4). The most important metal phthalocyanines are derived from phthalodinitrile, phthalic anhydride, Pc derivatives, or alkali metal Pc salts.

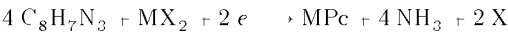
The route from *o*-phthalodinitrile [91-15-6] can be represented $4\text{ C}_8\text{H}_4\text{N}_2 + \text{M} \longrightarrow \text{MPc}$, where M is a bivalent metal, metal halide, metal alcoholate, or an equivalent amount of metal of valence other than two in a 4:1 molar ratio. If a solvent, eg, trichlorobenzene, benzophenol, pyridine, nitrobenzene, or quinoline, is used, the reaction takes place at approximately 180°C. Without a solvent the dry mixture must be heated to ca 300°C to initiate the exothermic reaction (50).

The following shows the reaction of phthalic anhydride [85-44-9] with urea:



In this process, catalysts, such as boric acid, molybdenum oxide, zirconium, and titanium tetrachloride or ammonium molybdate, are used to accelerate the reaction. The synthesis is either carried out in a solvent (aliphatic hydrocarbon, trichlorobenzene, quinoline, pyridine, glycols, or alcohols) at approximately 200°C or without a solvent at 300°C (51,52).

The synthesis from phthalimide derivatives, eg, diimidophthalamide (or phthalimide [85-41-6]) is usually carried out in a solvent such as formamide.



Metal phthalocyanines may also be prepared using alkali metal salts or from metal-free phthalocyanine by boiling the latter in quinoline with metal salt.

Industrial production of copper phthalocyanine usually favors either the phthalic anhydride–urea process (United States, United Kingdom) (1,52,53) or the *o*-phthalodinitrile process (Germany, Japan) (54,55). Both can be carried out continuously or batchwise in a solvent or bake process of the solid reactants (56).

The crude copper phthalocyanine must be treated to obtain a satisfactory pigment in regard to the crystal modification and optimal particle size (1,57) (see PIGMENTS). The particle size of crude phthalocyanine can be reduced by chemical or mechanical methods. The former involves dissolving phthalocyanine in concentrated sulfuric acid and precipitating it by addition to water (1). Alternatively, it may be slurried in dilute acid, washed, and dried, but this process leaves the pigment partially undissolved. A grinding step might have to be added because different sulfuric acid concentrations yield different sulfate levels. The α -modification, which is obtained by these methods, is treated by adding chlorine or sulfuric acid to the phthalocyanine, either during or after the synthesis to prevent recrystallization into the more stable β -form (57,58). This may be induced by heat or wet-milling in certain solvents (1,58). Very unstable modifications, like the reddish, chlorine-free α -copper phthalocyanine, can be stabilized with amides or salts of copper phthalocyanine sulfonic acids (59–63). Mixture with other metal phthalocyanines, eg, tin, vanadium, aluminum, or magnesium, also inhibits crystallization change and poor performance in binders and prints (floculation) due to the hydrophobic character of unsubstituted phthalocyanines.

The second process to finish phthalocyanine, which is more important for β -copper phthalocyanine, involves grinding the dry or aqueous form in a ball mill or a kneader (64). Agents such as sodium chloride, which have to be removed by boiling with water after the grinding, are used. Solvents like aromatic hydrocarbons, xylene, nitrobenzene or chlorobenzene, alcohols, ketones, or esters can be used (1). In the absence of a solvent, the crude β -phthalocyanine is converted to the α -form (57,65) and has to be treated with a solvent to regain the β -modification. The aggregate structure also has an impact on the dispersion behavior of α - and β -copper phthalocyanine pigments (66).

Incorporation of less than a stoichiometric amount of alkyl sulfonamides of copper phthalocyanines into copper phthalocyanine improves the pigment's properties in rotogravure inks (67). Monomeric and polymeric phthalocyanine derivatives with basic substituents adsorb strongly to the pigment surface and promote the adsorption of binder molecules (68–72).

Performance in ink and coatings can be improved by addition of surfactants (qv), dispersants, resins, or copper phthalocyanine derivatives with long aliphatic chains, $\text{CuPc}(\text{CH}_2\text{--NHR})_3$ (68,73), to stabilize the pigment in the binder system. Another possibility is wet-milling of aqueous pigment dispersions incorporating an organic medium, eg, glycols, polyethers, or surfactants (74).

γ -Copper phthalocyanine is obtained by treating the α -modification with 30% sulfuric acid and glycol monobutyl ether at 110°C (1,75,76). δ -Copper phthalocyanine is prepared by dissolving copper phthalocyanine in 98% sulfuric acid, adding water, benzene, and turkey red oil, washing with alcohol and water, and drying (77,78). ϵ -CuPc is made by mixing urea, sulfuric acid, and CuPc made from the phthalic anhydride–urea process (1,41).

Some references cover direct preparation of the different crystal modifications of phthalocyanines in pigment form from both the nitrile–urea and phthalic anhydride–urea process (79–85). Metal-free phthalocyanine can be manufactured by reaction of *o*-phthalodinitrile with sodium amylate and alcoholysis of the resulting disodium phthalocyanine (1). The phthalic anhydride–urea process can also be used (86,87). Other sodium compounds or an electrochemical process have been described (88). Production of the different crystal modifications has also been discussed (88–93).

Perchloro- and perchlorobromo copper phthalocyanine [1328-53-6] are important organic green pigments. They are accessible through direct chlorination of copper phthalocyanine in a eutectic melt of aluminum and sodium chloride or in a chlorosulfonic acid medium (94,95). Bromine can be used instead of chlorine in the $\text{AlCl}_3\text{--NaCl}$ melt to obtain polybromochloro copper phthalocyanine. Other synthesis paths have been disclosed (96–99). Solvents such as methanol (qv), butanol, ethylene glycol (see GLYCOLS), esters (100), ketones (101), or benzoic acid (qv) (102) are used as finishing agents to convert the halogen phthalocyanines into the finely divided pigment forms.

Phthalocyanine sulfonic acids, which can be used as direct cotton dyes (1), are obtained by heating the metal phthalocyanines in oleum. One to four sulfo groups can be introduced in the 4-position by varying concentration, temperature, and reaction time (103). Sulfonyl chlorides, which are important intermediates, can be prepared from chlorosulfonic acid and phthalocyanines (104). The positions of the sulfonyl chloride groups are the same as those of the sulfonic acids (103). Other derivatives, eg, chlormethyl phthalocyanines (105–107), *tert*-butyl (108–111), amino (112), ethers (109,110,113–116), thioethers (117,118), carboxyl acids (119–122), esters (123), cyanides (112,124–127), and nitrocompounds (126), can be synthesized.

Polymeric phthalocyanines, which possess a higher stability compared to the monomers, can be obtained by combining a phthalocyanine with a polymer (1). The linking of the polymeric chain can occur at the central metal atom, the phenyl rings, through bridging or attachment to a polymeric chain (14,17,127–131). Various phthalocyanine analogues, which are potential pigments, have been studied (1). For example, the four-fused benzene ring systems can be replaced by other aromatic rings like naphthalene (132–135) or other heterocyclic systems (136–152).

Specifications, Standards, Analysis, and Quality Control

Testing of phthalocyanines includes crystallization (qv), flocculation, and application in paints, plastics (qv), and printing inks (1). The ASTM standard specifications include CuPc in dry powder form for various applications (153). The specifications cover color (qv), character or tint, oil absorption, reactions in identification tests, and dispersions and storage stability. Quantitative determinations are possible with ceric sulfate (30) or sodium vanadate (154). Identification methods are given (155), including tests for different applications.

Uses

Approximately 90% of the phthalocyanines (predominantly copper phthalocyanine) are used as pigments (qv). In addition, they have found acceptance in many types of dyestuffs, eg, direct and reactive dyes, water-soluble and solvent-soluble dyes with physical and chemical binding, azo-reactive dyes, azo nonreactive dyes, sulfur dyes, and vat dyes (1) (see DYES; DYES, REACTIVE).

Available Forms. Phthalocyanines are available as powders, in paste, or liquid forms. They can be dispersed in various media suitable for aqueous, nonaqueous, or multipurpose systems, eg, polyethylene, polyamide, or nitrocellulose. Inert materials like clay, barium sulfate, calcium carbonates, or aluminum hydrate are the most common solid extenders. Predispersed concentrates of the pigments, like flushes, are interesting for manufacturers of paints and inks (156), who do not own grinding or dispersing equipment. Pigment–water pastes, ie, presscakes, containing 50–75% weight of water, are also available.

Colorants. The pigmentary forms of copper phthalocyanine are by far the most important commercial products of that class. They provide excellent color properties, excellent resistance to heat and light, acid and alkali, and are extremely insoluble in most solvents. They are less expensive than other organic pigments and color practically every type of printing ink, paint, plastic, and textile. Other uses include the coloring of roofing granules, cements and plasters, fine art paint, soaps, detergents, and other cleaning products (157). The two principal classes of copper phthalocyanine pigments are the blues and the greens. The blues may be further classified as the α - and β -crystal types, and the greens as the chlorinated and brominated derivatives.

Phthalocyanine Blues. α -Copper phthalocyanine blue is a reddish species used primarily in coatings and plastics. Several varieties are marketed. The basic form, the unstable Pigment Blue 15 [147-14-8] (CI 74160), is used in water-based paints, paints containing weak petroleum solvents, and in certain plastics, eg, PVC, that require mild processing conditions.

The unstable α -modification tends to crystallize in the presence of strong solvents or heat, which causes flocculation and a significant loss of color strength. Stabilized forms exist as the noncrystallizing (NC) types, eg, Pigment Blue 15:1, and the noncrystallizing, nonfloculating (NCNF) types, eg, Pigment Blue 15:2. Additives provide improved rheological properties and reduce flocculation. The NCNF types are stable to heat and a variety of solvents. Plastics and certain metal decorating inks are the main applications for the NC types, whereas the NCNF types are the preferred colorants for coatings, like industrial and automotive paints. Many multipurpose systems for architectural finishes, which come in contact with a variety of solvents and are used under varied conditions, are made, at least partly, of NCNF blues (see PAINT, ARCHITECTURAL).

β -Copper phthalocyanine blue (Pigment Blue 15:3) is characterized by the greenish hue and cleanliness of color tone, which is necessary for the cyan component in color process printing. It is, however, weaker in color strength than α -copper phthalocyanine blue. Used as the cyan in four-color process printing (yellow, magenta, cyan, black), all the spectral colors can be reproduced in the different printing systems, eg, letter press, gravure, flexographic, offset, screen-printing, and textile inks (see PRINTING PROCESSES). For applications incorporating strong solvents at higher temperatures (gravure printing), a nonfloculating (NF) type, Pigment Blue 15:4, is favorable. Relatively smaller amounts of β -copper phthalocyanine (compared to the α -modification) are used in coatings and plastics because of the lower color and strength. The color purity of β -CuPc is not required, since most systems for plastics and coatings are shaded with other pigments.

ϵ -Copper phthalocyanine blue 15:6 is significantly redder and slightly stronger in color strength than the alpha-crystal type. The ϵ -type can also be produced in a stabilized form, so that crystallization and flocculation are minimized in systems containing strong solvents (41), which makes applications in coatings and printing inks possible.

Metal-free copper phthalocyanine blue, ie, Pigment Blue 16 [574-93-6], is one of the earliest forms of phthalocyanine. Environmental concerns about copper in pigments tended to increase the use of metal-free copper phthalocyanine, but certain shortcomings (greenish hue, lack of stability in aromatic solvents) allowed only specialty uses (109). The stabilized α -NC-type is used in certain automotive coatings.

Phthalocyanine Green. The commercial development of the perchloro- and perchlorobromo copper phthalocyanine greens began in the 1930s. The predominant form of the halogenated derivatives of phthalocyanines is the chlorinated Pigment Green 7 [1328-53-6] (CI 74260). The blue hue shifts toward yellow when chlorine is replaced by bromine, eg, Pigment Green 36 [14302-13-7] (CI 74265). This form was first marketed in the late 1950s. Like nonhalogenated compounds, halogenated phthalocyanines must be converted to useful pigments by finishing in liquids, eg, methanol, butanol, or ethylene glycol. Phthalocyanine green pigments (158,159) are used in a variety of paints and plastics. Because the green color in the printing process is predominantly obtained by combinations of yellow and blue, the application in printing inks is reduced to spot colors. Fluorine-treated pigments have been used as writing inks (160).

Phthalocyanine Dyes. In addition to their use as pigments, the phthalocyanines have found widespread application as dyestuffs, eg, direct and reactive dyes, water-soluble dyes with physical or chemical binding, solvent-soluble dyes with physical or chemical binding, azo reactive dyes, azo nonreactive dyes, sulfur dyes, and wet dyes. The first phthalocyanine dyes were used in the early 1930s to dye textiles like cotton (qv). The water-soluble forms like sodium salts of copper phthalocyanine disulfonic acid, Direct Blue 86 [1330-38-7] (CI 74180), Direct Blue 87 [1330-39-8] (CI 74200), Acid Blue 249 [36485-85-5] (CI 74220), and their derivatives are used to dye natural and synthetic textiles (qv), paper, and leather (qv). The sodium salt of cobalt phthalocyanine, ie, Vat Blue 29 [1328-50-3] (CI 74140) is mostly applied to cellulose fibers (qv).

Sulfonamide groups can be introduced into phthalocyanines to make them soluble in alcohols and glycol ethers. Their main applications are transparent paints, flexo and gravure printing inks, wood stains, plastics, and ballpoint inks. This includes products like Solvent Blue 24 (CI 74380), Solvent Blue 42 [1330-38-7] (CI 74180), and Solvent Blues 44, 46, and 52.

Incorporating additional tertiary or secondary amine groups, the sulfonamide group leads to dyes which are soluble in acidic media, eg, printing inks, ballpoint ink, ink ribbons, copying inks, and carbon paper. In contrast to the phthalocyanine dyes already mentioned, developing dyes do not contain any substituents and possess purer, more genuine colors, eg, Ingrain Blue 2:1 (CI 74160) (161–163). Derivatives of copper phthalocyanine sulfonates with diphenylguanidine are used for ballpoint inks (164,165).

Other Uses. Phthalocyanines have interesting properties as catalysts, lasers (qv), semiconductors, lubricants, or as photographic components. Metal phthalocyanines catalyze dehydrogenations, oxidation, or gas-phase reactions. Catalysis by metal phthalocyanines was first discovered in 1936 in the exchange reaction between hydrogen and water and the catalysis of water from oxygen and hydrogen (166). The catalysis of oxygen transfer reactions suggested the similarity of metal phthalocyanines to heme and enzymes like hydrogenase (167). A variety of chemical compounds, eg, alkanes, olefins, aromatics, alcohols, aldehydes, alkyl aromatics, phenols, thiols, cumenes, polymers, and sugar, are oxidized by molecular oxygen when iron, copper, or cobalt phthalocyanine are present (1,43,44). For example, cobalt phthalocyanine sulfonate [30638-08-5] and vanadium phthalocyanine sulfonate [42862-24-8] catalyze air oxidation of mercaptans and other sulfur compounds of petroleum (qv) (1,168). They oxidize caustic solutions that are used to dissolve sulfur compounds from petroleum, hydrogen sulfide, and disulfides. Magnesium phthalocyanine [1661-03-6] and iron phthalocyanine [136-16-1] catalyze the oxidation of cumene by air (169). Cobalt phthalocyanine [3317-67-1] catalyzes the oxidation of liquid toluene and ethyl benzene to benzyl alcohol (170). Isopropyl alcohol has been oxidized to acetone with the use of copper or other metal phthalocyanines as a catalyst (171). In addition, hydrogenation, dehydrogenation, degradation, polymerization, isomerization, hydrogen-exchange reaction, reactive dehydrogenation, hydrogenative thermal cracking, autoxidation, epoxidation, decarboxylation, and Fischer-Tropsch synthesis are catalyzed by phthalocyanines (42–44). For example, a number of phthalocyanines are able to catalyze the hydrogenation of carbon monoxide to form C₁–C₅ hydrocarbons (172), the formation of ammonia at room temperature from hydrogen and nitrogen in contact with a sodium film (173) or the polymerization of methacrylates, nitriles (qv), and polymethylenes (1). Phthalocyanines have been extensively studied for application in fuel cells (qv) (45–47). Cobalt phthalocyanine especially was studied to enhance the oxygen release or to act as an electrode. Iron phthalocyanine has even higher electrocatalytic activity (48,174). Polymeric phthalocyanines exhibit even higher stability (175).

Air oxidation of dyestuff waste streams has been accomplished using cobalt phthalocyanine sulfonate catalysts (176). Aluminum has been colored with copper phthalocyanine sulfonate (177,178). Iron phthalocyanine can be used as a drier in wood oil and linseed oil paints (179).

Metal-free, chloroaluminum phthalocyanine [14154-42-8], vanadyl phthalocyanine [13930-88-6], or magnesium phthalocyanines are sufficiently soluble in organic solvents and show enough bleachable absorption at 694.3 nm to serve as repeated Q-switching elements for ruby lasers (qv) (180). Phthalocyanines have been used in other lasers as well (181).

Phthalocyanines have been used to incorporate semiconductor properties in polymers (182) or to develop a thin-film transistor (183). Phthalocyanines and their derivatives can act as dyes in color photography (qv) (184) or electrophotography (185). Light-sensitive compositions for use on lithographic plates are comprised in part of copper phthalocyanine blue (186). Dichlorosilicon phthalocyanine [19333-10-9] has been used in the

photoconducting layer in the target area of a television camera tube (187). Copper polyphthalocyanine can be prepared with an excess of copper which produces a material with *p*- or *n*-type conduction for use in photoelectric devices and solar cells (188). The conducting and photoconducting properties of metal-free phthalocyanine and of phthalocyanine containing less than the theoretical amount of copper have been studied (189). Doping of other semiconductors and general photoconducting properties of phthalocyanines have been disclosed (1,185,190).

Phthalocyanines are excellent lubricants at temperatures of 149–343°C (191). Combinations with other lubricants, like grease, molybdenum, or tungsten sulfides, have found applications in the automotive industry or professional drilling equipment (192–195). Further uses include indicators for iron(II), molybdenum(V), and uranium(IV) (196) or redox reactions (197), medical applications like hemoglobin replacements (198) or sterilization indicators (199), or uses like in gas filters for the removal of nitrogen oxides from cigarette smoke (200).

Health and Safety Factors

Phthalocyanines do not pose any significant risk to human health in the environment or the workplace. The blue and green pigments have LD₅₀ values of >10 g kg body weight (201). Skin and eye irritation studies on CuPc (202) and the Ames test for mutagenicity were negative (203). In several studies, no carcinogenic risk or toxicity to humans was revealed. The FDA approved the use of CuPc in general and ophthalmolic surgery, for contact lenses, and food packaging (qv) (204). Phthalocyanine Blue [147-14-8] may be used as a colorant for coatings that are used in manufacturing, packing, processing, preparing, treatment, packaging, transporting, or holding food (205). The TLV value for CuPc is 10 mg m³ (206). Other potential health hazards, which can be reduced to meet existing limits through protective equipment, have been discussed (207). A list of trace metals in typical commercial phthalocyanines has been given (208).

Polychlorinated biphenyls (PCBs) have been detected in pigments manufactured in trichlorobenzene, but not in those made with nonchlorinated solvents (209). High boiling hydrocarbons or esters are suitable replacements (210–212). Producers of copper phthalocyanine pigments may be faced with various regulations, based on the presence of copper in these pigments (201,207), although the U.S. EPA has delisted Pigment Blue 15 as a source of copper for reporting under SARA III, Section 313. OSHA has established PEL of 0.1 mg m³ for copper fumes (201). Other regulations deal with the issue of copper in waste water (EPA) (204) and solid wastes (RCRA) (201).

Economic Aspects

Phthalocyanine pigments account for approximately 23° of the total worldwide organic pigment consumption of 225,000 tons. Approximately 20,000 t are used in printing inks, 10,000 t in paints, 9,000 t in plastics, 3,000 t in textiles, 7,000 t in dyes, and 2,000 t in specialty uses. Table 1 shows the worldwide distribution of crude phthalocyanine capacity. The production history of phthalocyanine in the United States from 1980 to 1990 is given in Table 2 (161). The 1990 prices of phthalocyanine blue and green pigments were ca \$11–22 kg and \$21–27 kg, respectively.

Table 1. Worldwide Distribution of Crude Phthalocyanine Capacity

Country	Estimated capacity, t
United States, Canada	5,500
Western Europe	13,500
Eastern Europe	3,000
Japan, Taiwan, Korea, India	33,500
South America	1,500

Table 2. U.S. Production of Phthalocyanine Blue and Green Pigments, t

Colorant	1981	1983	1985	1987	1989	1991 ^a
Blue 15	412	631	354	564	195	221 ^b
Blue 15:1	476	429	452	432	380	503 ^b
Blue 15:2	548	357	258	229	198	199 ^b
Blue 15:3	3,720	3,956	4,196	5,752	8,017	8,460 ^b
Blue 15:4	^c	^c	580	661	644	609 ^b
<i>total</i>	<i>5,156</i>	<i>5,373</i>	<i>5,841</i>	<i>7,637</i>	<i>9,434</i>	<i>9,992^b</i>
Green 7	1,112	878	814	959	1,211	1,350
Green 36	99	89	66	43	42	38
<i>total</i>	<i>1,211</i>	<i>967</i>	<i>879</i>	<i>1,002</i>	<i>1,253</i>	<i>1,388</i>
<i>Total</i>	<i>6,367</i>	<i>6,340</i>	<i>6,720</i>	<i>8,639</i>	<i>10,687</i>	<i>11,388^b</i>

^a In 1991, no detailed figures were published for phthalocyanine blue, but the total volume dropped 3.3° compared to 1990.
^b Value is estimated.
^c No information available.

BIBLIOGRAPHY

"Phthalocyanine Dyes" in *ECT* 1st ed., Vol. 10, pp. 607–611, by K. C. Johnson, E. I. du Pont de Nemours & Co., Inc.; "Phthalocyanine Compounds" in *ECT* 2nd ed., Vol. 15, pp. 488–494, by F. H. Moser, Holland-Suco Color Co.; in *ECT* 3rd ed., Vol. 17, pp. 777–787, by F. H. Moser, Consultant, and W. H. Rhodes, BASF Wyandotte Corp.

- F. H. Moser and A. L. Thomas, *Phthalocyanine Compounds*, Reinhold Publishing Co., New York, 1963.
- A. B. P. Lever, in H. J. Emeleus and A. G. Sharpe, eds., *Advances in Inorganic Chemistry and Radiochemistry*, Vol. 7, Academic Press, Inc., New York, 1965, pp. 27–113.
- L. J. Boucher, in G. A. Melson, ed., *Coordination Chemistry of Macrocyclic Compounds*, Plenum Press, New York, 1979, Chapt. 7.
- F. Lux, in C. J. Kevane and T. Moeller, eds., *Proceedings of the 10th Rare Earth Research Conference 1973*, p. 871.
- K. Kasuga and M. Tsutsui, *Coord. Chem. Rev.* **32**, 67 (1980).
- P. Sayer, M. Goutermann, and C. R. Connell, *Acc. Chem. Res.* **15**, 73 (1982).
- C. Hamann and co-workers, in P. Goerdlich, ed., *Organische Festkoerper und duenne Schichten*, Akademische Verlagsgesellschaft, Leipzig, Germany, 1978, Chapt. 2.

8. M. S. Fischer, D. H. Templeton, A. Zalkin, and M. Calvin, *J. Am. Chem. Soc.* **93**, 2622 (1971).
9. F. Cariatì, F. Morazzoni, and M. Zocchi, *Inorg. Chim. Acta* **14**, L31 (1975).
10. J. R. Mooney, C. K. Choy, K. Knox, and M. Kenney, *J. Am. Chem. Soc.* **97**, 3033 (1975).
11. A. Gieren and W. Hoppe, *J. Chem. Soc., Chem. Commun.*, 413 (1971).
12. W. E. Bennet, D. E. Broberg, and N. C. Baenziger, *Inorg. Chem.* **12**, 930 (1973).
13. K. Kasuga and co-workers, *J. Am. Chem. Soc.* **102**, 4835 (1980).
14. D. Wuhde, *Adv. Poly. Sci.* **50**, 45 (1983).
15. T. Inabe, J. W. Lyding, and T. J. Marks, *J. Am. Chem. Soc., Chem. Commun.*, 1084 (1983).
16. C. W. Dirk, T. Inabe, K. Schoch, and T. J. Marks, *J. Am. Chem. Soc.* **105**, 1539 (1983).
17. E. A. Orthmann, V. Enkelmann, and G. Wegner, *Makromol. Chem. Rapid Commun.* **4**, 687 (1983).
18. O. Vulker, *Naturwiss. Rundsch.* **8**, 268 (1955).
19. F. M. Smith and J. D. Easton, *J. Oil Colour Chem. Assoc.* **49**, 614 (1966).
20. E. A. Lawton, *J. Phys. Chem.* **62**, 384 (1958).
21. Y. Taru and K. Takoaka, *Shikizai Kyokaishi* **55**(1,2) (1982).
22. M. Radulescu and R. Vilceanu, *J. Therm. Anal.* **7**(1), 209 (1975).
23. D. Horn and B. Honigmann, XII, *Fatipec. Congress Vlg. Chemie*, Weinheim, Germany, 1974, p. 181.
24. **Ger. Pat. 1,181,248** (Nov. 12, 1964), B. I. Knudsen and H. S. Rolskov (to Kemisk Vaerk Koge AS).
25. B. Honigmann and D. Horn, *SCI Monogr.* **38**, 283 (1973).
26. B. Honigmann and D. Horn, *Particle Growth in Suspensions*, Academic Press, Inc., New York, 1973.
27. M. Shigemitsu, *Bull. Chem. Soc. Jpn.* **32**, 607 (1959).
28. B. Honigmann, *Farbe Lack* **70**, 787 (1964).
29. **U.S. Pat. 3,357,989** (Dec. 12, 1967), J. F. Byrne and P. F. Kurz (to Xerox Corp.).
30. **Ger. DE-AS 2026057** (Dec. 17, 1970), P. J. Brach and H. A. Six (to Xerox Corp.).
31. H. A. Rutter and J. D. McQueen, *J. Inorg. Nucl. Chem.* **12**, 361 (1961).
32. J. A. E. Elvidge and A. B. P. Lever, *Proc. Chem. Soc. London*, 195 (1959).
33. G. E. Ficken and R. P. Linstead, *J. Chem. Soc.*, 4847 (1952).
34. **Brit. Pat. 322,169** (Nov. 18, 1929), A. G. Dandridge, H. A. Drescher, and J. Thomas (to Scottish Dyes Ltd.).
35. M. Shigemitsu, *Bull. Chem. Soc. Jpn.* **32**, 544 (1959).
36. M. A. Dahlen, *Ind. Eng. Chem.* **31**, 839 (1939).
37. **U.S. Pat. 2,613,128** (Oct. 7, 1952), F. Baumann and B. Bienert (to Farbenfabriken Bayer AG).
38. Ref. 35, p. 502.
39. **U.S. Pat. 2,613,128** (Oct. 7, 1952); **Brit. Pat. GB 704,310** (Feb. 17, 1954), B. Bienert and F. Baumann (to Bayer AG, Germany).
40. F. Grund, *J. Soc. Dyers Colour.* **69**, 671 (1953).
41. **Ger. Pat. DE-OS 2210072** (Sept. 6, 1973); **Brit. Patent GB 1,411,880** (Oct. 29, 1975); **Fr. Pat. FR 2,174,089** (Oct. 29, 1986), B. Honigmann and J. Kranz (to BASF AG).
42. H. Meier, *Organic Semiconductors*, Verlag Chemie, Weinheim, Germany, 1974.
43. S. Manassen, *Catal. Rev.* **9**(2223) (1974).
44. F. Steinbach, H. H. Schmidt, and M. Zobel, *Catal. Proc. Int. Symp.*, 417 (1975).
45. J. H. Wood, C. W. Keenan, and W. E. Bull, *Fundamentals of College Chemistry*, 3rd ed., Harper and Row, New York, 1972.
46. R. Jasinski, *J. Electrochem. Soc.* **112**, 526 (1965).
47. H. Jahnke, M. Schoenborn, and G. Zimmermann, *Proc. Symp. Electrocatal.*, 303 (1974).
48. J. Yamaki, *J. Appl. Electrochem.* **15**, 441 (1985).
49. O. Hirabaru and co-workers, *J. Chem. Soc., Chem. Commun.*, 481 (1983).
50. *P.B. 65687*, Bios Final Report no. 960, item 22.
51. **U.S. Pat. 2,197,458** (Apr. 16, 1940), M. Wyler (to Imperial Chemical Industries Ltd.).
52. **U.S. Pat. 3,188,318** (June 8, 1965), D. E. Mack (to Hercules Powder Co.).
53. **U.S. Pat. 4,173,658** (Nov. 6, 1979), R. Schmitz and K. Wunderlich (to Bayer AG).
54. **Ger. Pat. DE-OS 1,644,679** (Nov. 19, 1970), E. Spieltschka and W. Deucker (to Hoechst AG).
55. **U.S. Pat. 3,985,767** (Oct. 12, 1976), J. C. R. Nicaise and L. A. Cabut (to Produits Chimiques Ugine Kuhlmann).
56. Y. L. Meltzer, *Chem. Process. Rev.* **42** (1970).
57. F. M. Smith and J. D. Easton, *J. Oil. Colour Chem. Assoc.* **49**, 614 (1966).
58. **Eur. Pat. 574,792** (Dec. 22, 1993), E. Dietz and M. Urban (to Hoechst HG).
59. V. A. Smrcek, S. G. Vilner, V. P. Gorin, and G. A. Burgrov, *Lakokras. Mater. Ikh. Primen.* **2**, 11 (1984).
60. **Ger. Pat. DE-OS 2,720,464** (Nov. 24, 1977), I. R. Wheeler and G. H. Robertson (to Ciba-Geigy).
61. **Ger. Pat. DE-OS 2,461,567** (July 17, 1975), G. Hull and A. Topham (to ICI).
62. **Ger. Pat. DE 2,519,753** (Nov. 11, 1976), **DE 2,516,054** (Nov. 11, 1976), W. Jettmar, B. Honigmann, and R. Sappok (to BASF AG).
63. **Brit. Pat. GB 1,422,834** (Jan. 28, 1976), H. Leister (to Bayer AG).
64. **Brit. Pat. GB 470,079** (Aug. 9, 1937), G. W. Johnson (to IG Farbenind.).
65. **U.S. Pat. 2,791,589** (May 7, 1957), A. Pugin (to Ciba-Geigy).
66. H. Benkreira and C. M. Davie, *Colloids Surf. A* **90**(1), 37 (1994).
67. **U.S. Pat. 3,754,958** (Aug. 28, 1973), V. A. Giambalvo (to American Cyanamid Co.).
68. **Brit. Pat. GB 949,739** (Feb. 19, 1964), **GB 985,620** (Mar. 10, 1965), A. Schoellig, R. Schroedel, and H. J. Sasse (to BASF AG).
69. **Brit. Pat. GB 1,082,945** (Sept. 13, 1967), **GB 972805** (Oct. 14, 1964), W. Black, J. Mitchell, and A. Topham (to ICI).
70. W. Black, F. T. Hesselink, and A. Topham, *Kolloid Z. Z. Polym.* **213**, 150 (1966).
71. A. Topham, *Prog. Org. Coat.* **5**, 237 (1977).
72. J. Moilliet and D. A. Plant, *J. Oil Colour Chem. Assoc.* **52**, 289 (1969).
73. Jpn. Koka: Tokkyo Koho, **03247667** (Nov. 5, 1991), K. Sawamura and M. Hayashi (to Toyo Ink Mfg. Co., Japan).
74. **U.S. Pat. 4,853,039** (Aug. 1, 1989), T. E. Donegan, J. H. Bantjes, and T. G. Leary (to BASF Corp.).
75. **Rus. Pat. 670,970** (Jan. 19, 1967), Y. Abe (to Dainichiseika Colour Chem.).
76. S. Suzuki, Y. Bansho, and Y. Tanabe, *Kogyo Kagaku Zasshi* **72**(3), 720 (1969).
77. **U.S. Pat. 3,150,150** (Sept. 22, 1964); **Brit. Pat. GB 912,526** (Dec. 12, 1962), B. P. Brand (to Imperial Chemical Industries Ltd.).
78. **Rus. Pat. 670870** (Jan. 18, 1967), Y. Abe (to Dainichiseika Colour Chem.).
79. **Rus. Pat. 70 06,708** (Mar. 7, 1970), S. Horiguchi, Y. Abe, and T. Hosoda (to Dainichiseika Color and Chemicals Manufacturing Co., Ltd.).
80. **Rus. Pat. 70 06,709** (Mar. 7, 1970), S. Horiguchi, Y. Abe, and T. Hosoda (to Dainichiseika Color and Chemicals Manufacturing Co., Ltd.).
81. **Rus. Pat. 71 43,634** (Dec. 24, 1971), S. Ohira, T. Hosoda, and H. Otohà (to Dainichiseika Color and Chemicals Manufacturing Co., Ltd.).
82. **Rus. Pat. 71 43,636** (Dec. 24, 1971), Y. Abe and T. Hosoda (to Dainichiseika Color and Chemicals Manufacturing Co., Ltd.).
83. **Ger. Offen. 2,250,938** (May 2, 1974), M. Gaeng (to BASF AG).
84. **Ger. Offen. 2,255,117** (June 28, 1973), E. E. Jaffe (to E. I. du Pont de Nemours & Co., Inc.).

85. U.S. Pat. **4,152,171** (May 1, 1979), R. Barraclough and R. Langley (to Ciba-geigy Corp.).

86. U.S. Pat. **2,820,796** (Jan. 21, 1958), F. F. Ehrich (to du Pont).

87. U.S. Pat. **200,005,152** (May 7, 1935), J. F. Thorpe and J. Thomas (to ICI).

88. Ger. Pat. DE **2711005** (Sept. 21, 1978), P. Thoma, W. Habermann, and J. Kranz (to BASF).

89. U.S. Pat. **3,060,189** (Oct. 23, 1962); Brit. Pat. GB **952,775** (Mar. 18, 1964); Ger. Pat. DE **1,218,639** (Dec. 29, 1966), E. Stocker and A. Pugin (to J. R. Geigy AG).

90. Ger. Pat. DE-OS **2,364,689** (Sept. 5, 1974); Brit. Pat. GB **1,410,310** (1974); Fr. Pat. FR **2,212,399** (1974); Jpn.-Kokai **75157417** (1974); Czech. Pat. **579131** (1974); U.S. Pat. **3,998,839** (1976), Z. Seha (to Ciba-Geigy).

91. Brit. Pat. GB **998,225** (July 14, 1965); Neth. Pat. **6405836** (Dec. 2, 1964), H. Mueller (to BASF).

92. Eur. Pat. EP **592,955** (Apr. 20, 1994), P. Boettcher, P. Erk, and J. Jesse (to BASF AG).

93. Jpn.-Kokai **57/64158** (Oct. 8, 1982), R. Ohshima (to Toyo Ink Mfg. KK.).

94. U.S. Pat. **2,276,860** (Mar. 17, 1942), G. Hiemann, W. Schmidt, F. Muhlbauer, and G. Weist (to General Aniline and Film Corp.).

95. Ger. Offen. **2,052,937** (May 10, 1972), K. G. Kossenko, D. W. Tito, and A. P. Hopmeier (to Sun Chemical Corp.).

96. Ger. Offen. **2,637,891** (Mar. 10, 1977), M. A. Kleist and R. G. W. McKay (to Imperial Chemical Industries Ltd.).

97. Ger. Offen. **2,707,849** (Sept. 1, 1977), R. G. W. McKay (to Imperial Chemical Industries, Ltd.).

98. U.S. Pat. **4,077,974** (Mar. 7, 1978), D. Wessling (to Bayer AG).

99. U.S. Pat. **4,091,028** (May 23, 1978), R. Barraclough and R. Langley (to Ciba-Geigy AG).

100. U.S. Appl. **236,955** (1981), U.S. Appl. **275,488** (1981); Eur. Pat. EP-A **58888** (Sept. 1, 1982), J. F. Santimauro (to Mobay Corp.).

101. Ger. Pat. DE **3,636,428** (May 5, 1988), J. Kranz and K. Schmeidl (to BASF AG).

102. Eur. Pat. EP-A **0182207** (Oct. 16, 1991), J. Kranz (to BASF AG).

103. R. P. Linstead and F. T. Weiss, *J. Chem. Soc.*, 2977 (1950).

104. Ger. Pat. DE **891121** (Sept. 24, 1953), F. Nadler, H. Hoyer, and O. Bayer (to IG Farbenind.).

105. Brit. Pat. BE **565,782** (Sept. 17, 1958), E. Parette and F. Maes (to Ciba).

106. Brit. Pat. GB **586,340** (Mar. 14, 1947), N. H. Haddock and C. Wood (to ICI).

107. Ger. Pat. DE **843,726** (July 10, 1952), A. Tattter (to BASF AG).

108. M. J. Camenzind and C. L. Hill, *J. Heterocycl. Chem.* **22**, 2 (1985).

109. M. Hanack, J. Metz, and G. Pawloski, *Chem. Ber.* **115**, 2836 (1982).

110. J. Metz, O. Schneider, and M. Hanack, *Inorg. Chem.* **23**, 1065 (1984).

111. A. B. P. Lever and co-workers, *J. Am. Chem. Soc.* **103**, 6800 (1981).

112. D. Woehde and B. Schulte, *Makromol. Chem.* **186**(11), 2229 (1985).

113. N. O. Sigl, *J. Heterocycl. Chem.* **18**, 1613 (1981).

114. T. M. Keller and J. R. Griffith, *J. Fluorine Chem.* **12**, 73 (1978).

115. C. L. Leznoff and T. W. Hall, *Tetrahedron Lett.*, 3023 (1982).

116. C. Piechocki and co-workers, *J. Am. Chem. Soc.* **104**, 5245 (1982).

117. A. W. Snow and N. L. Jarvis, *J. Am. Chem. Soc.* **106**, 4707 (1984).

118. A. W. Snow, J. R. Griffith, and N. P. Marullo, *Macromolecules* **17**, 1614 (1984).

119. H. Shirai, A. Marayuma, K. Kobayashi, and N. Hoyo, *Macromol. Chem.* **181**, 575 (1980).

120. S. Higaki, K. Hanabusa, H. Shirai, and N. Hojo, *Macromol. Chem.* **184**, 691 (1983).

121. J. H. Schutten and J. Zwart, *J. Mol. Catal.* **5**, 109 (1979).

122. B. N. Achar, G. M. Fohlen, J. A. Parker, and J. Keshavayya, *Indian J. Chem. Sect. A*. **27A**(5), 411 (1988).

123. H. Shirai and co-workers, *Makromol. Chem.* **185**(12), 2537 (1984).

124. D. Wuhde, G. Meyer, and B. Wahl, *Macromol. Chem.* **181**, 2127 (1980).

125. D. Wuhde and B. Wahl, *Tetrahedron Lett.* **3**, 227 (1979).

126. Ger. Pat. DE **4,216,120** (Nov. 18, 1993), G. Beck and B. Kaletta (Bayer HG).

127. T. J. Marks, *J. Coat. Technol.* **48**, 53 (1976).

128. A. Meller and A. Ossko, *Monatsh. Chem.* **103**, 150 (1972).

129. K. Fischer and M. Hanack, *Chem. Ber.* **116**, 1860 (1983).

130. J. Metz, G. Pawlowski, and M. Hanack, *Z. Naturforsch.* **38b**, 378 (1983).

131. A. W. Snow and J. R. Griffith, in J. I. Kroschwitz, ed., *Encyclopedia of Polymer Science and Engineering*, 2nd ed., Vol. 11, John Wiley & Sons, Inc., New York, 1988, pp. 211–225.

132. S. A. Mikhalenko and E. A. Luk'yanets, *Zh. Obshch. Khim.* **39**(11), 2554 (1969).

133. T. A. Shatskaya, M. G. Gal'pern, V. R. Skvarchenko, and E. A. Luk'yanets, *Zh. Obshch. Khim.* **57**(10), 2364 (1987).

134. S. A. Mikhalenko, L. A. Yogodina, and E. A. Luk'yanets, *Zh. Obshch. Khim.* **46**(7), 1598 (1976).

135. Brit. Pat. GB **457,526** (Nov. 30, 1936), A. V. Aerden (to IG Farbenind.).

136. Ger. Pat. DE **696590** (Sept. 25, 1940), R. J. Recidon (to IG Farbenind.).

137. Brit. Pat. GB **471,418** (Aug. 30, 1937), W. W. Groves (to IG Farbenind.).

138. R. P. Linstead, E. G. Noble, and J. M. Wright, *J. Chem. Soc.*, 911 (1937).

139. A. R. Koray, V. Ahsen, and O. Bekaroglu, *J. Chem. Soc., Chem. Commun.* **12**, 932 (1986).

140. V. N. Kopranenkov and E. A. Luk'yanets, *Zh. Obshch. Khim.* **41**(10), 2341 (1971).

141. D. Wuhde, G. Schnurpfeil, and G. Knothe, *Dyes Pigm.* **18**(2), 91 (1992).

142. R. Hendriks, Ot E. Sielcken, W. Drength, and R. J. M. Roeland, *J. Chem. Soc., Chem. Commun.* **19**, 1464 (1986).

143. N. Kobayashi and Y. Nishiyama, *J. Chem. Soc., Chem. Commun.* **19**, 1462 (1986).

144. M. Whalley, P. M. Brown, and D. B. Spiers, *J. Chem. Soc.*, 2882 (1957).

145. M. E. Baguley, H. France, R. P. Linstead, and M. Whalley, *J. Chem. Soc.*, 3521 (1955).

146. U.S. Pat. **2,850,505** (Sept. 2, 1958), D. W. Heinz (to American Cyanamid Co.).

147. U.S. Pats. **2,681,344** (June 15, 1954), 2,681,345 (June 15, 1954), H. Frances (to ICI).

148. Ger. Pat. DE **663,552** (Aug. 9, 1938) (to IG Farbenind.).

149. Rus. Pat. **59/00633** (Jan. 7, 1984) (to Sumitomo Chem. Co.).

150. Brit. Pat. GB **750,240** (June 13, 1956); U.S. Pats. **2,744,913** (May 8, 1956) **2,681,341** (June 15, 1954), W. F. Beech and co-workers (to ICI).

151. Ger. Pats. DE **964,324** (May 23, 1957), DE **1001785** (Aug. 22, 1957), G. Roesch, H. Klappert, and W. Wolff (to Bayer AG).

152. Ger. Pat. DE-AS **1,017,720** (Apr. 10, 1958), F. Erbe (Hoechst).

153. *1980 Annual Book of ASTM Standards*, Part 28, American Society for Testing and Materials, Philadelphia, Pa., 1980.

154. G. G. Rao and T. P. Sastri, *Z. Anal. Chem.*, 11 (1959).

155. R. L. P. R. Hepsworth, *Chem. Ind.*, 272 (1952).

156. T. A. Langstroth, in T. C. Patton, ed., *The Pigment Handbook*, Vol. 3, John Wiley & Sons, Inc., New York, 1973, p. 447.

157. F. H. Moser, in Ref. 156, Vol. 1, p. 684.

158. Jpn. Kokai Tokyo Koho **06 009916** (Jan. 18, 1994), H. Myashita and I. Konuki (to Pentel KK, Japan).

159. Ger. Pat. DD **301,408** (Dec. 24, 1992), K. R. Groth, F. Kleine, E. Schurig, and W. Dassler (to Chemie AG Bitterfeld-Wolfer, Germany).

160.

Jpn. Kokai Tokkyo Koho **06 080924** (Mar. 22, 1994), H. Idokawe, M. Wakata, N. Watanabe, and T. Nobuatsu).

161.

Synthetic Organic Chemicals: United States Production and Sales, 1980–1991, United States International Trade Commission.

162.

J. Eibl, *Melliand Textilber. Int.* **56**, 398 (1975).

163.

H. Vollmann, in K. Venkatamaran, ed., *The Chemistry of Synthetic Dyes*, Vol. V, Academic Press, Inc., New York, 1971, p. 283.

164.

U.S. Pat. 3,010,970 (July 29, 1959), N. W. Fiess (to American Cyanamid Co.).

165.

U.S. Pat. 3,470,200 (Sept. 30, 1969), N. L. Anderson (to Allied Chemical Corp.).

166.

M. Calvin, E. G. Cockbain, and M. Polanyi, *Trans. Far. Soc.* **32**, 1436 (1936).

167.

D. Rittenberg and A. I. Krasna, *Discuss. Faraday Soc.* **20**, 185 (1955).

168.

Brit. Pat. GB 849,998 (Sept. 28, 1960) (to Universal Oil Products).

169.

Ger. Pat. 1,041,960 (Oct. 30, 1958) (to Ruhrchemie AG).

170.

U.S. Pat. 3,256,324 (June 14, 1966), P. S. Landis, D. D. Neiswender, and R. D. Offenhauer (to Socony-Mobil Oil Co., Inc.).

171.

F. Steinbach and H. Schmidt, *J. Catal.* **29**(3), 515 (1973).

172.

M. Ichikawa, M. Sudo, M. Soma, T. Onishi, and K. Tamaru, *J. Am. Chem. Soc.* **91**(6), 1538 (1969).

173.

Ger. Offen. **2,006,347** (Aug. 6, 1970), M. Sudo and co-workers (to Tokyo University).

174.

J. Yamaki, *J. Electrochem. Soc.* **132**, 2125 (1985).

175.

K. Beales, D. D. Eley, D. J. Hazeldine, and T. F. Palmer, in H. Kropf and F. Steinbach, eds., *Katalyse an Phthalocyaninen*, Thieme, Stuttgart, Germany, 1973, p. 1.

176.

V. F. Borodkin, T. A. Ananeva, M. I. Alyanov, and G. S. Zueva, *Izv. Vyssh. Uchebn. Zaved. Khim. Khim Tekhnol.* **18**(8), 1286 (1975).

177.

U.S. Pat. 3,205,550 (Aug. 24, 1965), H. Grossman and H. Keller (to Durand and Hugenin AG).

178.

U.S. Pat. 3,326,728 (June 20, 1967), M. A. Robinson (to Olin Mathieson Chemical Corp.).

179.

H. P. Kaufmann and M. Bernard, *Fette Seifen Anstrichm.* **66**(8), 613 (1964).

180.

P. P. Sorokin, J. J. Luzzi, J. R. Lankard, and G. D. Pettit, *IBM J. Res. Develop.* **8**(2), 182 (1964).

181.

A. C. Seldon, *Brit. J. Appl. Phys.* **18**(6), 743 (1967).

182.

M. Becker and H. F. Mark, *Angew. Chem.* **73**, 641 (1961).

183.

S. E. Harrison and co-workers, NASA Accession No. N65-24791, Rep. No. AD 461315, National Aeronautics and Space Administration, Washington, D.C., 1965.

184.

Brit. Pat. 873,054 (Jan. 21, 1959), N. H. Haddock and E. MacDonald (to Ilford Ltd.).

185.

M. Smith and J. W. Weigl, in J. H. Dessauer and H. E. Clark, eds., *Xerography and Related Processes*, Focal Press, New York, 1965, Chapt. 6 and index.

186.

Ger. Offen. **2,506,200** (Aug. 21, 1975), N. Nishikawa, K. Kobayashi, and Y. Aotani (to Fuji Photo Film).

187.

H. Meier and W. Albrecht, *Ber. Bunsenges. Phys. Chem.* **73**(1), 86 (1969).

188.

U.S. Pat. 3,009,981 (June 1, 1959), B. S. Wildi and A. S. Epstein (to Monsanto Chemical Co.).

189.

T. Kenichi and co-workers, *Denki Tsushin Daigaku Gakubu Rikogaku Hen* **13**, 85 (1962).

190.

H. Meier, W. Albrecht, and U. Tschirwitz, *Ver. Bunsenges. Phys. Chem.* **73**(8–9), 795 (1969).

191.

U.S. Pat. 3,023,164 (Feb. 27, 1962), E. A. Lawton, C. W. Allen, and S. Cosgrove (to Battelle Memorial Institute).

192.

Russ. Pat. 139,391 (1961), G. Ya Entin and R. I. Nigokaya.

193.

U.S. Pat. 3,432,432 (Mar. 11, 1968), J. L. Derher (to Chevron Research Co.).

194.

J. E. Houghton, *U.S. Atomic Energy Comm.*, RD B N-1186, 1968, 23 pp.

195.

U.S. Pat. 3,265,617 (Aug. 9, 1966), C. M. Allen (to the Battelle Development Corp.).

196.

G. Rao and N. V. Rao, *Anal. Chem.* **190**, 213 (1962); **196**, 192 (1963).

197.

H. S. Gowda and R. Shakunthala, *Chem. Anal.* **55**(1), 30, 1966.

198.

Fr. Pat. M-2181 (Jan. 2, 1964), P. L. Barral.

199.

Czech. Pat. 159,163 (July 15, 1975), V. Rambousek and D. Krupa.

200.

U.S. Pat. 3,724,469 (Apr. 3, 1973), J. W. Reynolds and C. H. Sloan (to Eastman Kodak Co.).

201.

R. E. Gosselin, H. C. Hodge, R. P. Smith, and M. N. Gleason, *Chemical Toxicology of Commercial Products*, 4th ed., Williams and Wilkins, Baltimore, Md., 1976.

202.

NPIRI Raw Materials Data Handbook, Vol. 4, Lehigh University, Bethlehem, Pa., 1984.

203.

P. Miloy and K. Kay, *J. Toxicol. Environ. Health* **4**, 31 (1978).

204.

Ambient Water Quality Criteria for Copper, Publ. 85227023 U.S. Environmental Protection Agency, Washington, D.C., 1984.

205.

Fed. Reg. **42**(50), 14547 (Mar. 15, 1977).

206.

American Conference of Governmental Industrial Hygienists, 5th ed., Cincinnati, Ohio, 1986.

207.

P. G. Webb, *Am. Ink Maker* **65**(2), 11 (1987).

208.

Am. Inkmaker **51**, 31 (Oct. 1973).

209.

R. C. Buchta and co-workers, *J. Chromatog.* **325**(2), 456 (1985).

210.

Jpn.-Kokai **52/036130** (Mar. 19, 1977), Y. Abe (to Dainichiseika Colour Chem.).

211.

Ger. Pat. DE-OS 2,642,416 (Mar. 31, 1977), A. Sato and co-workers (to Toyo Ink MFG KK).

212.

Jpn.-Kokai **52010326** (Jan. 26, 1977), T. Funazo (to Toyo Ink MFG KK).

General References

G. Booth, in K. Venkataraman, ed., *The Chemistry of Synthetic Dyes*, Vol. V, Academic Press, Inc., New York, 1971, p. 241.

C. C. Leznoff and A. B. P. Lever, eds., *Phthalocyanines: Properties and Applications*, VCH Verlagsgesellschaft, Weinheim, Germany, Vol. 1, 1989; Vols. 2 and 3, 1993.

Gerd Loebbert
 BASF Corporation

PICKLING OF STEEL.

See METAL SURFACE TREATMENTS.

PIGMENT DISPERSIONS

A pigment dispersion in a concentrated form is a uniform distribution of very fine color pigment particles in a suitable medium or carrier. Such a dispersion is normally used for applying color to the surface of a substrate, such as an ink film on paper or a paint film on a steel surface. It is also used for mass coloring, as in the case of plastics. Considering the high cost and specialized equipment in its preparation, a dispersion is manufactured in relatively small batches in highest concentration of pigment. The concentrate made in such a manner is usually diluted, reduced, or extended to produce the finished product.

Dispersion

Organic and inorganic pigment powders are finely divided crystalline solids that are essentially insoluble in application media such as ink or paint (1). The carrier used for dispersion of a pigment is usually a liquid or solid, such as a polymer, that is deformable at the processing conditions of high temperature and or shear. The color strength of the dispersed pigment increases markedly with decrease in particle size. Optimum color strength from a given pigment in practice requires a mean particle size of the order of 0.1 μm or less, which is half the wavelength of the light involved (2). Therefore, the dispersion process involves size reduction of the pigment particle to the smallest practical size, reasonably complete wetting of its solid surfaces by the carrier, and stabilization of the resulting dispersion.

Because the intensity and color strength of pigments are largely dependent on the exposed surface, it is desirable to reduce the particles to primary particle size. This is the size of the solid pigment crystals as they are precipitated in their synthesis. In practice, the size reduction processes are limited by the nature of pigment, dispersion system, constraints of the processing equipment, the requirements imposed by the product application, and the overall economics. The maximum aggregate size permissible in a given dispersion system depends on the thickness of the film or the coating. For example, the dispersion used for architectural coatings can tolerate a much larger pigment aggregate than a similar dispersion used for automotive finishes, which requires finer particles. Any dispersion system, however, is expected to contain a very small number of these largest aggregates. Generally, it is important to reduce most aggregates to the smaller size to achieve color strength, gloss, film integrity, and durability.

In a dispersed pigment system, a primary pigment particle refers to an individual crystal and a loosely formed association of the pigment crystals from the manufacturing process. Size reduction beyond primary particle size requires excessive energy, but it also has an adverse effect on the visual properties of the pigment. Generally, the particle size of most organic pigments is much smaller initially by precipitation than optimum primary particles, but the particles tend to grow to a much larger size when their formation is complete (3–6).

Organic pigments, such as the azo red and yellow pigments, in the process of striking the color undergo definite crystal growth following their precipitation from the aqueous media (see AZO DYES). The individual crystals are joined together due to forces on the crystal surfaces to form the aggregate. These are held together as static systems by van der Waals forces. Subsequent processing to recover the pigment product results in the formation of agglomerates, which are large associations of pigment crystals and aggregates. The agglomerates are held together by forces that are much weaker than those present within the aggregates. Typically, agglomerates are joined at the edges and comers in a loose matrix form. It is possible to generate an even larger association of pigment agglomerates or flocculates during further processing. These formations are loosely held together and are usually easy to break down by application of shear. Various surface treatments are used to suppress the formation of large aggregates and, thereby, ease the dispersion process. These treatments range from the classical approach of rosination to additions of a variety of surface-active agents at the synthesis step (4,7). However, occasionally large agglomerates, several millimeters in diameter, form during the initial stages of dispersion in a highly viscous system (8). The commercial processes used in dispersion manufacturing may not fully eliminate the aggregates. However, the design and operation of pigment dispersion equipment is aimed at application of mechanical forces to break down the agglomerates and even some less tightly held aggregates. Ideally, an excellent dispersion should consist mainly of primary pigment particles and few loosely held aggregates (9).

Wetting of the pigment surface constitutes a critical step in achieving a stable and uniform pigment dispersion. Wetting refers to displacement of adsorbed gases (usually air) on the surface of pigment particles, followed by attachment of a vehicle system to the pigment surface. Since the vehicles used for many dispersion systems are viscous, it follows that the penetration of vehicles to the pigment surface is slow and, hence, aided by external mechanical forces. Thus, the grinding (size reduction) and wetting of pigment are frequently carried out simultaneously. The adsorbed gases are displaced on application of shear, and the action also provides smearing of vehicle on the pigment surface and exposes a new surface for wetting. The system of wetted fine primary pigment particles must be stabilized to prevent reversal of the dispersion process. It is usually done by surrounding the particles with a protective colloid or buffer which blocks the reagglomeration action of particles. In some cases, the stabilization is attained by addition of ions to establish similar charges on all particles.

Flushing

Flushing processes are used extensively in preparing organic pigment dispersion concentrates for color printing ink applications. The process can be described as a direct transfer of pigment from an aqueous phase to an oil or nonaqueous phase without drying. When the pigment presscake is mixed with an oil-based vehicle or a carrier, water is separated from the pigment surface and replaced by the vehicle. Most organic pigments demonstrate an affinity for hydrocarbon oils and lend themselves to easy dispersion in oil by the process of flushing (see PIGMENTS, ORGANIC). Inorganic pigments, on the other hand, have to be treated with cationic surfactants to make their surface lipophilic (10). The majority of inorganic pigments are usually dried and dispersed as dry powders in the carrier, as opposed to being flushed. Techniques used for dispersion of these pigments are different and should be treated as special cases (7,10–13) (see PIGMENTS, INORGANIC).

The process of flushing typically consists of the following sequence: phase transfer; separation of aqueous phase; vacuum dehydration of water trapped in the dispersed phase; dispersion of the pigment in the oil phase by continued application of shear; thinning the heavy mass by addition of one or more vehicles to reduce the viscosity of dispersion; and standardization of the finished dispersion to adjust the color and rheological properties to match the quality to the previously established standard.

The equipment used for the flushing process is a heavy-duty sigma blade dough mixer, generally referred to as a flusher in the organic pigment industry. The mixer is equipped with two nonoverlapping blades which can be operated at different speeds, if desired. The mixer is usually jacketed to permit heating or cooling of the mass inside. It is a heavy-duty and expensive machine, requires a high energy input, and its operation is labor intensive.

Typically, the power input ranges from W kg (0.25–0.75 horsepower units gal) of processed mass. Hydraulic drives, mechanical, and electronic controls are also used to achieve variation in speed of the blades for efficient processing.

When the pigment in the water phase and vehicle are mixed, the oil begins to emulsify, increasing the interfacial area. The pigment at the oil–water interphase transfers to the oil phase. As the pigment particles in the oil phase coalesce, water from the highly viscous mass begins to separate from the mixture. Ideally, the separated water is free of oil and or pigment particles when the phase transfer is complete. The water so separated is poured by tilting the mixer hydraulically. Although 80–90% of water is removed mechanically by the process of phase transfer, a small amount of water is still present in the pigment–oil dispersion in the emulsified and trapped form. Complete removal of water from the dispersion is achieved by applying heat and or vacuum. The process is somewhat difficult during the first step, as there is a chance that the water could emulsify with the oil phase. The clean separation of water in the first step is important to achieve the high quality of dispersion. The mechanical separation of water also removes the salts present in the pigment filter cake. Occasionally, an additional amount of salt present in the dispersion is removed after the phase-transfer step by leaching the mass with repeated additions of water. The mechanism of flushing is well established in the industry and has been described in various publications (15,16).

Flushing is frequently used for the manufacture of large quantities of a dispersion having a specific pigment in a compatible vehicle system. The flushed products, typically containing 28–40% pigment, offer sufficient flexibility to the formulator to produce the finished offset ink. The flushed products exhibit superior gloss, transparency, and strengths, compared to those produced by dispersing the dried pigment. Flushing is particularly important to dispersions of organic pigments, such as Diarylide Yellow (CI Pigment Yellow 12, CI 21090) and Alkali Blue (Pigment Blue 61, CI 42765) because the drying process is detrimental to the product quality of these pigments.

Equipment

Various types of equipment are used commercially to manufacture dispersed pigment concentrates or finished dispersion products used by printing ink, and the coatings and plastics industry.

Kneaders or Internal Mixers. Kneaders are designed to process materials and mixtures with high plastic viscosity, ie, up to 10 Pa·s (100 P). Dispersion of pigment is accomplished by kneading the process mass containing pigment, water, and vehicle. The mechanical energy is transferred throughout the high viscosity mass by internally developed shear. Typically, the mixer is equipped with two parallel helical blades rotating in the opposite direction at different speeds in a contoured vessel. The blades and the mixer are machined to a high degree of precision to keep the clearance between the blades and between the blade and the vessel wall as low as possible. The wiping action of the blades produces a high degree of shear. The mass continuously undergoes the action of folding and unfolding by the opposing pitch of the mixing blades. The system formulation is controlled to develop a high plastic viscosity, which allows transfer of mechanical energy between the moving blade and the vessel wall. The bulk of the water separates from the mass as a clean aqueous phase and is poured off by tilting the mixer hydraulically. Small amounts of remaining water are removed under vacuum, leaving behind a highly viscous mass of pigment in the vehicle. Considerable heat is generated during the process, and it is removed partially by circulating water in the jacket and adjusting the viscosity of the mass, by addition of thinners and oils, to regulate the energy input.

The heavy-duty sigma blade batch mixer is used widely for the process of flushing in the preparation of pigment dispersions, primarily for offset inks. For some products, it is also used to disperse dry pigments in liquid vehicles of a wide range of viscosity and in thermoplastic resins with relatively low melting points. The finished dispersion from the flusher can be discharged as a solid or liquid with a wide range of viscosities. The mixer offers a wide operating range of pressures, temperatures, and speeds of blades. This processing flexibility makes it an extremely versatile machine for processing a wide range of dispersions and explains its popularity for producing pigment dispersions for offset inks, despite its high installation and operating costs. The mixers are routinely available to 3774-L (1000-gal) capacity and can process up to 3200 kg of finished dispersion. The choice of construction material ranges from carbon steel and chrome-plated carbon steel to various stainless steel alloys. Often the mixers are hydraulically driven to operate under constant torque and permit operation with variable speeds at different stages. Due to very specialized construction, close tolerances, and large drive motors, the installations require high capital investments. The process is batch-type and can take several hours to complete. Considering the high capital and operating costs, alternative continuous processing systems have been explored to produce similar dispersions. There are, however, few successful commercial installations.

A two-roll mill can be considered as an internal mixer due to a large clearance (up to 15 mm) between its rolls. The process mass is subjected to kneading just before it enters the nip between the rolls. Two-roll mills are used to prepare pigment dispersion in elastomers and thermoplastic resins. In some cases, the pigment and the medium is premixed; in others, the resin is banded on one roll and the pigment is added. This type of equipment provides good control of the plastic viscosity, resulting in high quality dispersions. The rolls are parallel and rotate in opposite directions. The degree of shear can be varied by varying the roll speeds, the distance between rolls, and the temperature. The operating temperature is controlled by circulating water or steam through a core of rolls. Electric resistance heaters or hot oil circulation is used if very high temperatures are required. Two-roll mills are heavily powered to subject a small mass to high energy input to achieve the desired dispersion in a short cycle time, typically a few minutes. The equipment is used to produce small quantities of concentrates and, as such, has limited commercial applications (17).

Close Tolerance Mills. This class of dispersion equipment can be classified into low speed cylindrical roll mills and high speed disk or cone mills. Cylindrical roll mills are used for processing pigment dispersions in the paste form for ink and paint applications. Processing viscosities are relatively moderate, compared to the mass processed in flusher, and range from 20–1000 Pa·s (200–10,000 P). The cylindrical roll mills comprise multiple, parallel rotating rolls powered by a single drive motor. As many as five adjacent rolls rotating in opposite direction to each other have been used, although three rolls are most common. A good premix of pigment and vehicle is fed to the nip of the slowest moving rolls and removed by a doctor blade from the fastest moving rolls. The clearance between the rolls is typically 10–15 μm, requiring high precision in the manufacture of the rolls. The clearance between the rolls decreases successively from the feed end to the product end. This decreasing nip clearance and the increasing rotational speeds of the subsequent rolls exert very high shear on the material passing through the nip. The inward rotation of the rolls forces the process mass toward the nip, and the bulk of it is forced back up and continually dragged back toward the nip. This flow pattern creates a high degree of mixing and shearing. Mathematical expressions have been developed for horsepower requirements, throughput, and applied work, and have practical applications in operation of the mills. The rolls are cored for circulation of cooling water or steam. The construction of the rolls is robust, with the roll surface machined to a high degree of precision so that it runs parallel with unchanging gap over the entire length of the roll. The rolls are machined to have a slight degree of convexity (crown) to account for operation at temperatures of 60°C and higher and under high pressure.

High speed stone and colloid mills achieve dispersion by smearing action. The equipment consists of two accurately shaped carborundum stones, one stationary and the other rotating at high speed (3600–5400 rpm) with a small gap separating the stones. The shear applied to the mass flowing between the stones under laminar conditions provides the dispersing action. Although the colloid mill is designed primarily to produce emulsion-type colloidal dispersion, it is also used for preparation of pigment dispersions. The rotors are in the form of disks, or truncated cones, rotating at 3000–5000 rpm. The dispersion is fed by gravity, and often the flow in the machine resulting from centrifugal force is aided by the feed pump. The viscosities are typically moderate at 10 Pa·s (100 P) or less. The energy input depends on the gap (clearance between stator and rotor), surface roughness, and rheological properties of the process material (17).

High Speed Fluid Energy Mills. This type of equipment is used primarily for preparation of relatively low viscosity mill bases for inks and paints. The first is an impeller type which achieves dispersion by the application of shear. The second type is in the form of a rotor–stator, and the dispersion is achieved by impingement or impact.

The high speed disk disperser consists essentially of a circular saw blade-type impeller mounted vertically in a cylindrical tank. The rotational speed of the impeller is very high with a common peripheral (tip) speed of 5000 ft min (18). Various impeller designs are available, depending on the type, size, and direction of angle of the serration along the rim. Frequently, a dual shaft agitator system, consisting of a low speed anchor agitator mounted centrally in the tank and the high speed disperser mounted off-center, is used to provide good mixing and dispersion. Since the dispersion is achieved by shearing action, the mixer becomes effective at high viscosities, such that a sufficient flow is maintained in the tank with a flow pattern resembling a rolling doughnut within the tank. The high viscosity is usually obtained by increasing the pigment solids in the dispersion, rather than increasing the vehicle

viscosity. Also, the typical application involves mixing the dry pigment in the vehicle, and wetting the pigment surface by penetration, leading to choice of vehicles with relatively low viscosity. Considering the mild dispersing action of this impeller, compared to the other equipment described previously, such as flushers, the use is limited to easily dispersed pigments. A lot of empirical work has been done to provide the guidelines for formulating dispersions, specifically for high speed disk disperser blades. The disk disperser is also used to prepare a premix or a mill base, which is processed further to obtain a finished product.

The high speed dispersion mill is based on impingement action. A kinetic dispersion mill, eg, the Kady mill manufactured by Kinetic Dispersion Corp., is a typical example of this type of equipment. The mixer is a rotor–stator combination mounted vertically in a cylindrical tank. The slotted stator fits like a collar around the slotted rotor with clearance between the stator and rotor of less than 2 mm. The peripheral velocity (tip speed) of the rotor is very high, typically ranging from 40–50 m/s. The pigment–vehicle mixture is drawn in from the top and bottom of the mill head and forced through the slots in the rotor–stator by centrifugal force. The break-up and attrition of the pigment aggregates occur as a result of forceful impact. The suspended solids emerging from the rotor slots at high velocity are subjected to additional impact against the stator wall and against each other. Because the disintegration of large particles occurs by impingement rather than by shear, the formulations using low viscosity vehicles can be used, exploiting more effective penetration and melting properties of such vehicles. The vehicle solids can be added after the dispersion is complete to achieve its stabilization.

Ball and Pebble Mills. A ball or pebble mill is essentially a cylindrical container partially filled with metallic or ceramic grinding media in the form of balls or pebbles. The mill base, consisting of pigment and vehicle, is charged to the mill. Grinding and dispersion takes place because of impact and shear resulting from the cascading action of the balls when the cylinder is rotated around its horizontal axis. The ball mills and the pebble mills are differentiated on the basis of the nature of grinding media and the liner. Ball mills typically use metallic media (cast iron or alloy steel), and the liner is constructed out of alloy steel or some other special material. On the other hand, pebble mills utilize ceramic media and are lined with a nonmetallic liner, such as ceramic or rubber. The efficiency of the ball mill depends on a number of physical factors, such as size and speed of the mill; size, density, and loading of the media; and load, viscosity, and composition of the mill base. The mill is typically operated at a speed which induces a cascading action of the balls. The optimum angular velocity is a function of the critical speed, the speed at which a complete centrifugation occurs. The optimum speed is $20.35/\sqrt{r} - 1.75/\sqrt{r}$ where r is radius in meters.

The loading of grinding media is typically 40–50% of the mill volume, whereas the mill base is about 20%. The dispersion technology in the ball mill is fairly well understood, and high quality products are routinely manufactured using the ball or pebble mills in the paint and ink industry (19). It is a versatile piece of equipment suitable to carry out premixing and dispersion steps in a single machine. The maintenance costs are usually low, and although it is batch processing equipment, it requires virtually no supervision to operate. It also offers wide latitude in formulations, from the standpoint of viscosity and pigment loading, and can handle pigments that are hard to disperse. The main disadvantage of this type of equipment is that it is rather bulky, and considering its long batch cycle time, has limited throughput. Ball and pebble mills have dominated in the dispersion of paints and inks over the past 50 years. The trend has been to change the manufacturing processes from the ball mills to the continuous processing systems, based on media mill, because of their higher throughput and ability to process high viscosity bases.

Some modifications in the basic concept of the ball mill has led to the development of attritors and vibration mills (20,21). The attritor consists of a vertical grinding chamber equipped with an agitator. The agitator basically includes a set of horizontal fingers attached to the central shaft and provides the necessary action for dispersion by vigorously agitating the grinding media. The equipment can be operated as a batch or continuous processing system. The continuous processing system is equipped with a recirculating pump for the mill base. It is claimed that an attritor can reduce the cycle time significantly, compared to the ball mills. The higher efficiency can be attributed to the fact that all of the grinding media is in constant motion, unlike the media in the ball mill. Typically, the media used in the attritors is 0.3175–0.476 cm, which is smaller than that used in the ball mill.

Sand, Bead, and Shot Mills. Sand, bead, and shot mills have replaced the ball and pebble mills in many applications. This class of equipment can be broadly termed as media mills. They have gained in popularity largely because they require less space, are capable of operating continuously, have higher specific energy input, and consequently have much higher throughput when compared to a ball mill with comparable capital investment. They can be regarded as a logical extension of ball (pebble) mills and then the attritors. Generally, the media mill consists of a cylindrical, jacketed chamber fitted with a high speed agitator. With the exception of the sand mill, the media mill can be positioned vertically or horizontally. The chambers are filled anywhere from 65 to 90% of the gross volume with grinding media. The earliest version of this type of equipment was a sand mill, which utilized fine 840–590 μm (20–30 mesh) Ottawa sand. The dispersion or mill base is introduced at the bottom with a product overflow from the top. The separation of the media from the product is aided by a screen in the outlet. The intense agitation of the fine media provides multiple contacts for the pigment agglomerates, and size reduction and dispersion of the solid takes place by a combination of smear and impact. The processing temperatures can be controlled by use of a heating or cooling medium. The operation of sand mills is limited to atmospheric pressure; hence mill bases with volatile solvents, or those at high viscosity, cannot be handled.

The bead and shot mills have overcome some of the limitations inherent with the design of sand mills. The bead and shot mills are similar in construction and differ only in the fact that bead mills use ceramic or glass media, varying in size from 0.3 to 3 mm, whereas shot mills use similarly sized carbon steel and alloy steel shots in addition to ceramic media. Depending on the application, a horizontal or vertical type of mill is used. The vertical mills are simpler and robust in design, exhibit higher throughput, and are better suited for a wide range of viscosity bases. The media mills in horizontal and vertical configuration have been extensively used in the paint and ink industry for dispersion of pigments. The horizontal mill offers the benefits of uniform distribution of media, ease of starting with large chamber volume, and ease of serviceability. The chamber volume ranges from 0.25 L for a laboratory size mill to as large as 500 L for large-scale industrial installations. The agitator is configured in the form of a series of flat disks with holes or a series of pins. The tip speed varies between 12–20 m/s. The separation of media from the dispersion is accomplished by cylindrical sieve cartridges or rotating gap separators. Application of the gap separator is usually limited to a media size of 1.00 mm or greater. The slot width (gap) has to be less than half of the average bead diameter. The trend has been toward bead sizes of ≤ 0.5 mm to achieve the dispersion quality needed. The quality of dispersion and the specific energy input is influenced by the size, hardness, and density of the media, tip speed, media loading, and viscosity of the mill base (22). Advances in media mill technology include use of a milling chamber in the form of a narrow annular space allowing extremely high specific energy input (23,24) and novel designs of media separators (22). The quality and production costs of the dispersed products manufactured using the media mills have improved significantly in the 1990s, but capital costs of the sophisticated installation has also risen steadily. A significant amount of work has been done to develop a mathematical model of the process in the media mill as applicable to the dispersion of pigments (25).

Uses

The formulation of dispersed pigment concentrates is influenced by the manufacturing process, as well as the performance parameters desired in the final application. The finished product in many cases is significantly different in formulation than the concentrate to achieve desired properties. One of the principal factors to be considered is the concentration of pigment in the dispersion concentrate. Compatibility of the carrier (solvent additives, etc) used in the preparation of concentrated dispersion and that used in the finished color product also plays an important role. In some cases this can be difficult because the carriers having the best performance, from the standpoint of processing, could be poor in the application systems. However, in the majority of the applications, particularly in coatings and colored plastics, the concentration of the pigment in the finished product is quite low, and the incompatibility problem is easily overcome.

Generally, the pigment dispersion concentrates are formulated for specific end use. They can be supplied as flushed pigments, dispersions or pastes for offset inks, chip dispersions for solvent and aqueous inks, and color concentrates for coloring large quantities of plastics. Although it is feasible for the end user to prepare the pigment dispersion concentrates, it is usually more cost effective and technologically advantageous to manufacture these dispersions by the pigment manufacturers or specialty dispersion houses. Three significant areas of application for concentrated dispersions, ie, printing inks, coatings, and plastics, are considered in the following.

Printing Inks. The consumption of dispersed pigment concentrates in the form of flushed color pastes is substantial in the manufacture of offset printing inks (qv). The film thickness of the ink film on the substrate (usually paper) ranges from 0.002 to 0.01 mm. The concentration of pigment in a typical ink formulation is relatively high at 8–20%, depending on the type of pigment and end use. Therefore, dispersed pigment concentrates

are formulated at very high pigment level to permit the flexibility in formulation of the printing ink products. For example, the flushed color concentrates used in offset printing inks are formulated with pigment concentration ranging from 20 to 50%. These are extremely heavy viscous pastes with viscosities ranging from 100 to 10,000 Pa·s (1000–100,000 P). In some applications, the dispersed pigment concentrate is in the solid (chip) form, containing 40–70% pigment in a carrier resin, such as nitrocellulose. These are converted to finished inks by dissolving the solid concentrate in the solvent, rather than by thinning or dilution.

The high pigment level in the concentrates permits use of high shear to obtain quality dispersions. Moreover, it allows greater economy for distribution of the concentrated products to the ink manufacturers. Most manufacturers of pigment concentrates or flushes in the United States offer a wide variety of products in liquid, paste, or solid forms, containing the specific pigment in the carrier or solvent compatible with most common types of inks, ie, letterpress, lithographic or offset, solvent gravure, and flexographic (packaging) inks. The estimated market for printing ink colorants, other than carbon black, extenders, and titanium dioxide, was 41.3×10^6 kg, valued at \$501.1 million. The estimated 1991 consumption of colorants in these four types of inks by form and type of ink is given in Table 1 (26).

Table 1. Estimated 1991 Consumption of Printing Ink Colorants by Form and Type of Ink, 10⁶ kg

Type of ink	Dry ^a	Flushed ^b	Resin bonded ^c	Solvent dispersion (dry basis)	Chip	Presscake ^d	Total
letterpress	0.14	0.91	0.0	0.0	0.0	0.0	1.1
lithographic	0.86	16.5	0.0	0.0	0.0	0.0	17.4
gravure	5.09	0.18	0.27	1.05	2.27	0.86	9.7
flexographic	2.0	0.0	0.45	0.64	0.27	7.0	10.4
other	1.54	0.23	0.59	0.0	0.36	0.0	2.7
Total	9.63	17.8	1.31	1.7	2.91	7.9	41.3

^a Includes easily dispersed pigments.
^b Including fluorescent pigments for lithographic inks.
^c Includes fluorescent pigments for gravure, flexo, and screen inks.
^d Includes aqueous dispersions.

In the United States, the bulk of color oil inks, ie, letterpress and lithographic, are manufactured from dispersed organic pigment concentrates, usually in the flushed paste form. The flushed colors are usually manufactured by the pigment manufacturers and supplied to the printing ink producers for conversion to the finished form. These flushes are formulated specifically for a particular application. For newspaper inks, where the ink film setting occurs mainly by penetration of oil in the papers, the pigment flushes use predominantly mineral oil and a small amount of resin for wetting the pigment. The pigment flushes for lithographic ink application (heatset and quickset) are formulated with a specific resin system and carrier solvent to permit printing with the least amount of emulsification. Lithographic inks, based on setting of film by application of heat, are prepared from the flushes that typically contain synthetic or modified natural resins dissolved in hydrocarbon oil (bp = 220–280°) and other specific additives. The ink film is cured by a combination of setting the resin system and evaporation of hydrocarbon oil. Flushes used in the manufacture of quickset inks are formulated with pigments in the carrier system containing synthetic resin, drying oils, and low viscosity nondrying oils. The film is initially set by penetration of low viscosity oil in the substrate, followed by oxidative polymerization of the resin and drying oils (qv) (27,28).

Gravure inks, typically used for long press runs, are very fluid (low in viscosity) and dry by evaporation of solvent, leaving behind an ink film ca 10-μm thick. Gravure inks are manufactured by milling dry pigment in the carrier system, consisting of volatile aromatic solvent, such as toluene, binders, and synthetic resins. Inks made by this procedure typically follow the sequence of shot milling the concentrate at a high pigment level ranging from 25–35%, followed by thinning the concentrate with the addition of solvent, resin solution, and other additives. Gravure inks are also manufactured from dispersed pigment concentrates in the form of resin chips by dissolving them in the carrier solvent system. The bulk of gravure inks, however, are manufactured with dry pigment as a starting material (27,28).

Flexographic inks are similar to gravure inks in that they are formulated as resin–solvent–pigment systems of low viscosity, and they dry by evaporation of solvent. The flexographic printing process differs from gravure printing in that it is based on the use of rubber rollers and plates and restricts the choice of solvents to low boiling alcohols and water in combination with oxygenated solvents. Dispersed pigment concentrates are used, similarly to gravure systems, except for differences in solvents and resins, which are principally acrylic polyamide, alcohol soluble nitrocellulose, and shellac (28,29). Concentrates in the form of solid chips are available in the market for aqueous inks suitable for textile printing. The chip typically contains 60% organic pigment and 40% of water-soluble acrylic resin. The chip is solubilized in an aqueous color base that can be combined with various compatible textile print clears to produce a finished textile ink (30).

Outlook. Total 1991 U.S. ink consumption was estimated at about 86×10^6 kg valued at over \$3.0 billion. The demand is estimated to grow at an average of about 3–4% per year in volume (26). The principal changes expected in the 1990s will continue to be in response to environmental and safety concerns and government regulations. The bulk of printing inks use raw materials based on fossil fuels, such as coal (qv) and petroleum (qv). Consequently, their cost and availability are an important factor influencing the direction in which the industry heads. Manufacturers of dispersed pigments must make appropriate changes in their product formulations to meet the demands of lithographic and letterpress ink businesses. The emission of volatile organic chemicals (VOCs) in the press room has increasingly led formulators of lithographic inks to innovative formulations. Also, the quicksetting ink applications use of irradiation is increasing to accelerate the oxidative polymerization process of film setting (29).

The flush colors for heatset inks are formulated with decreasing quantities of deodorized solvent or mineral ink oils, in response to the trend toward higher solids concentration in the ink, and consequently lower volatile compounds. There is a growing market for flushed colors containing soybean oil replacing a portion of mineral oil in the formulation (29,31). The benefits (eg, low emissions) of soya-based inks, use of renewable resources, and some enhancement in color properties have long been available in products with other vegetable oils, such as linseed or tung oils, both of which have been available for decades.

Although safety and environmental concerns have encouraged manufacturers to develop formulations based on aqueous pigment concentrates, such inks have not been widely accepted in high quality solvent gravure printing processes. The use of aqueous inks is largely limited to flexographic printing processes.

Coatings. Coatings (qv) generally exhibit heavier films, a fraction to a few millimeters thick, and are significantly lower in concentration of pigment. Dispersed pigment concentrates are therefore lower in pigment concentration and easier to reduce to the final product, when compared to similar dispersions prepared for offset inks. The concentrates are formulated to meet rather demanding performance specifications of the coated systems. The superior performance required for cooling systems is due to expected durability in (frequently) severe environments.

The use of water-based coating systems has grown steadily during the 1980s due to growing environmental concerns, but also because of continued improvement in the performance of such systems. Several additives are required in aqueous dispersions for coatings, including the dispersants tailored for the type of pigment used. Emulsifiers and thickeners to prevent settling of solids and preservatives to prevent bacterial growth during storage are also needed. The concentrates range from 20–40% pigment. Frequently, concentrates and the finished products are manufactured by the same dispersion house and many times at the same location.

Plastics. An estimated 12–16% of total U.S. organic pigment goes into these markets. Organic pigments, because of their generally poor heat stability, find somewhat limited use, but a few organic colorants can withstand high processing temperatures and have adequate lightfastness and bleed resistance for application in plastics (see COLORANTS FOR PLASTICS). Among them are diarylide yellows, phthalocyanine blue and green, permanent red 2B,

quinacridone, and perylene (32). Pigment dispersions, commonly known as color concentrates, are used in the plastic industry and show a wide variation in the pigment concentration and physical form. Color concentrates in various forms, ranging from liquids to solid granules, are metered directly in the processing equipment, usually by an extruder to incorporate color into the plastic process mass. The solid concentrates, however, are premixed with clear plastic prior to feeding the processing equipment, where both are melted and intensely mixed. Although the pigment concentration in the final product varies, depending on the nature of the application, it is usually quite low (<1% by weight), except in the case of very thin opaque plastic films. Extruders are typically used for reduction of color concentrates.

The color concentrates are manufactured, either by incorporating dry pigment in a compatible resin system in a high intensity mixer, or by the classical flushing process, wherein the pigment presscake is flushed with a low melting resin, followed by cryogenic grinding of the solid mass in the kneader (33). The latter process is claimed to be superior for pigments that are harder to disperse by conventional processes. Typically, the pigment content of the concentrate is between 10–50%. The resin used as a carrier needs to fulfill two principal requirements. It should have good wetting characteristics for the pigment for which it is used. Secondly, the carrier, which is usually a low melting thermoplastic, should be compatible for the thermoplastic polymer for which the concentrate is intended. A number of other dispersants and additives are used for the treatment of pigments to develop superior dispersion properties (34,35). Color concentrates in the pellet form are used most widely for coloring thermoplastics, particularly low density polyethylene (LDPE). The addition of pigment in high concentration changes properties of the carrier resin significantly. Specifically, concentrates become more difficult to melt because of the reinforcing effect of the pigment. Ideally, the melting point of concentrates should be similar to that of the unpigmented polymer. Hence, resin with a somewhat lower melting point than the target resin is chosen as a carrier resin for color concentrate.

The choice of a pigment for a specific plastic application also depends on its resistance to solvent and its insolubility in the polymer used. The phenomenon of migration, which includes the effects of bleeding and blooming, results from the partial dissolution of pigment in the polymer system at the processing temperature. Organic pigments generally satisfy the lightfastness requirement, but it is possible that a certain pigment may fade badly on exposure to light, although it is perfectly stable in the other dispersion systems.

Organic pigments are selected for applications, such as films and fibers, in which transparency and high tinctorial strength are desirable. Inorganic pigments, on the other hand, are preferred for those applications in which hiding power or opacity and high light- and weather-fastness are critical factors. Frequently, the two types are combined to take advantage of each economically. The following discusses the primary types of plastics with regard to preparation of pigment concentrates.

Thermoplastics. The highest consumption of color concentrates is in thermoplastic resins, such as low and high density polyethylene, polypropylene, PVC, and polystyrene. Processing techniques for thermoplastics are usually based on dry color dispersion in a compatible resin (36).

A wide variety of color concentrates are available for coloring rigid and flexible (plasticized) poly(vinyl) chlorides (PVC). Color concentrates for rigid PVC are made by dispersing a pigment in the resin. The flexible vinyls, on the other hand, use either concentrates made in plasticized resins or in a plasticizer, such as dioctyl phthalate (DOP) which is a liquid. The dispersions in the plasticizer are produced as a paste on a three-roll mill. The pigment concentration in the paste is typically 20–35%. The solid color concentrates for PVC are produced from free-flowing granules to a very fine powder (see VINYL POLYMERS, VINYL CHLORIDE AND PVC).

Polyolefins are manufactured and used in much greater quantity than any other class of plastics. The principal polyolefins are polyethylenes of various densities (LDPE, LLDPE, HDPE) and polypropylene (PP) (see OLEFIN POLYMERS).

The processing temperatures range from 160–260°C for LDPE to 220–300°C for PP. As a result of high processing temperatures, and consequently a high degree of softening, the shear force available during processing for the dispersion of pigment powder in polymer melt is limited. Therefore, pigment concentrates prepared in compatible polyolefins are frequently used. Color concentrates in the granulated form not only exhibit superior tinctorial strength, but also improve the safety of the operation. A number of mixing and processing techniques are used to manufacture the concentrates. They are manufactured in high and low density resins in internal mixers, two-roll mills, or sometimes in extruders. Dispersions in low molecular weight (polyethylene) grades are made by flushing techniques. Concentrates for film need the highest degree of dispersion, whereas those for injection molded articles may sacrifice dispersion quality for economy. The selection of pigment is done on the basis of its heat stability and tendency (or lack thereof) to migrate. Polypropylene color concentrates are more difficult to produce, due to its higher processing temperature and sharper melting points resulting from its crystalline structure.

Polystyrene (PS) is a highly rigid, glass clear, and almost colorless thermoplastic resin which softens between 80–100°C, and has a typical processing temperature range of 170–280°C. It is relatively easy to color; a significant volume is colored directly from pigment rather than dispersion concentrates. An adhesion or wetting agent is frequently used in conjunction with organic pigments for coloring polystyrene. Impact-resistant polystyrene containing typically 5–20% natural rubber is used widely for house appliances and products requiring good impact resistance. These are usually colored using the color concentrates. The concentrate manufacturing techniques are similar to those used for polyolefin. Copolymers of polystyrene with acrylonitrile and butadiene (ABS) are highly opaque, whereas polystyrene and styrene–acrylonitrile (SAN) are highly transparent. The choice of pigment used for coloring also depends on the opacity of polymer to be colored. Color concentrates for a polystyrene type in the paste form afford the ease of metering and blending various colors to obtain the required shade. Their applicability is somewhat compromised by the fact that their liquid component affects the mechanical properties of the finished colored plastic.

Dispersed pigment concentrates are available in a wide variety of other thermoplastic resins. Thermoplastic polyurethane (TPU) is used by the same organic pigments used for plasticized PVC. Dispersed pigment concentrates in carrier resins such as vinyl chloride–vinyl acetate copolymers, low molecular weight polyethylene, and PU itself are also used. Concentrates in the paste form, with pigment content ranging from 20–40%, are prepared by dispersing dry pigment powder in the PU solution in a ball or media mill. Other commercially available thermoplastics include polycarbonate, polyamides, poly(ethylene terphthalate) (PET), and cellulose derivatives. Pigmentation of these thermoplastics largely follows the same procedures as described above.

Thermosetting Plastics. Thermosets are processed only once, using heat and pressure to form semifinished or finished articles. The coloration of these plastics is generally accomplished using paste color dispersions. The colorants are worked into liquid resins before curing. Ball mills are commonly used to color prewetted molding powders prior to hardening. Dispersions for coloring epoxies are required to be free of water because the presence of water affects the curing role and hardeners of the resin. The dark color of phenolic resins restricts the coloration of these resins. Pigment dispersions used for thermosetting plastics are typically inorganic browns, reds, and greens. Coloration of unsaturated polyester and multiacrylic resin is performed with pigment–plasticizer (DOP) pastes. Occasionally, the pigments are directly dispersed into a small amount of monomer.

Coloration of Synthetic Fibers. Techniques of mass coloration or spin drying of synthetic fibers, particularly polyolefins, lie between the textile and plastic area. It is distinct from textile dyeing methods in that the material is colored before it is extruded. Pigment dispersion concentrates are widely used, considering their superior quality of dispersion when compared to incorporation of dry pigment powders, even if their average particle size is 2–3 μm. Large particles cause filament breakage and clogging of the spinnerette dies. As in the case of thermoplastic resins, the heat stability of the pigments used is critical. Spinning temperatures as high as 300°C used in melt spinning limit the use of only certain pigments. Pigment concentrates are made in the compatible polymer by the techniques described earlier. Polyolefin fibers, particularly polypropylene, represent the largest market for dispersed pigment concentrates (see FIBERS, OLEFIN FIBERS).

BIBLIOGRAPHY

"Pigments (Pulp and Flush Colors)" in *ECT* 1st ed., Vol. 10, pp. 690–695, by J. R. Briedy, Childs Pulp Colors Co., Inc.; "Pigments (Dispersed Concentrates)" in *ECT* 2nd ed., Vol. 15, pp. 589–605, by J. J. Singer, Holland-Suco Color Co.; "Pigments (Dispersed)" *ECT* 3rd ed., Vol. 17, pp. 871–889, by J. Singer, BASF Wyandotte Corp.

1. R. B. McKay and F. M. Smith, *Dispersion of Powders in Liquid*, 3rd ed., G. D. Parfitt, ed., Applied Science Publication, London, 1981.
2. R. B. McKay, *J. Oil Colour Chem. Assoc.* **71**(1), 7–10 (1988).
3. R. Sappok and B. Honigmann, in G. D. Parfitt and K. S. W. King, eds., *Characterization of Powder Surfaces*, Academic Press, Inc., New York, 1976.

4. B. Hays, *Amer. Ink Maker*, 28 (Nov. 90).

5. P. A. Lewis, ed., *Pigment Handbook*, Vol. 1, 2nd ed., John Wiley and Sons, Inc., New York, 1988.

6. D. D. Easton, *Can. Paint Finish.*, 20 (June 1975).

7. W. A. Kurtz, *Amer. Ink Maker*, 21 (June 1987).

8. W. H. Hobach, *Off. Dig. Fed. Paint Varn. Prod. Clubs*, 23(316), 255 (1951).

9. V. T. Crowl, *J. Oil Colour Chem. Assoc.* 55, 388 (1972).

10. *The Atlas HLB System*, 2nd ed., rev., Atlas Chemical Industries Inc., Wilmington, Del.

11. J. M. Rackman, *Pigm. Res. Technol.* 5, 5 (Apr. 1976).

12. *Ibid.* 10, 20 (May 1976).

13. *Ibid.* 5, 20 (June 1976).

14. *Ibid.* 11 (July 1976).

15. A. Mercado, *Amer. Ink Maker*, 36–41 (July 1991).

16. *Ibid.*, 54–59 (Aug. 1991).

17. T. C. Patton, *Paint Flow and Pigment Dispersion*, 2nd ed., John Wiley and Sons, Inc., New York, 1979.

18. Technical data, *Fundamentals of High Speed Dispersion*, Myers Engineering Inc., Bell, Calif.

19. Technical data, Paul O Abbe Inc., Little Falls, N.J.

20. Technical data, Sweco, Inc., Florence, Ky.

21. Technical data, Union Process, Inc., Akron, Ohio.

22. Technical data, Netzsch, Inc., Exton, Pa.

23. Technical data, Draisewerke, Inc., Allendale, Pa.

24. Technical data, Fryma, Inc., Edison, N.J.

25. J. Winkler, E. Klinke, and M. N. Satyanarayana, *J. Coating Technol.* 59(754), 35–60 (1987).

26. *Market Research Study U.S. Printing Ink Market, 1991–1996*, Hull & Co., Greenwich, Conn., 1992.

27. *Printer's Handbook*, Mclean Hunter Ltd., Barnet, 1992.

28. J. M. Fetsko, ed., *NPIRI Raw Materials Handbook*, Vol. 4, National Association of Printing Ink Manufacturers, Harrison, N.Y., 1983.

29. G. H. Hutchinson, *J. Oil Colour Chem. Assoc.* 72(7), 265–272 (1989).

30. S. George, *Amer. Ink Maker*, 67(6), 54 (1989).

31. A. Bean, *Amer. Ink Maker*, 65(5), 20, 47 (1987).

32. S. M. Rovito, M. K. Peters, L. S. Fujiise, A. Leder, J. Shimosato, and N. Takei "CEH Marketing Report—Pigments," *Chemical Economics Handbook*, SRI International, Menlo Park, Calif., 1990.

33. S. George, *Amer. Ink Maker*, 34 (June 1988).

34. A. C. D. Cowlely and R. Slater, *Plastics Eng.* 41 (Feb. 1989).

35. J. D. Bower, *Plastics Eng.* 33 (July 1984).

36. W. Herbst and K. Hunger, *Industrial Organic Pigments*, VCH, Weinheim, Germany, 1993.

General Reference

T. Patton, ed., *Pigment Handbook*, Vols. I, II, and III, John Wiley and Sons, Inc., New York.

Gerd Loebbert
Anand S. G. Sharangpani
BASF Corporation